Pathogenic camouflage as defense mechanism against autophagic recognition

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina

zu Braunschweig

zur Erlangung des Grades

eines Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Stefan Leupold
aus Dresden

1. Referent: Professor Dr. Dieter Jahn
2. Referent: Professor Dr. Wulf Blankenfeldt
eingereicht am: 30.09.2015
mündliche Prüfung (Disputation) am: 04.02.2016

Druckjahr 2016

Vorveröffentlichungen der Dissertation

Teilergebnisse dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Leupold, S., Völler, P., van den Heuvel, J. & Scrima, A.: The pathogenic camouflage mechanism of *Shigella flexneri*. (Vortrag) 16th Heart of Europe Bio-Crystallography Meeting, Hipping, Österreich (September 2013).

Leupold, S., Völler, P., van den Heuvel, J. & Scrima, A.: Pathogenic camouflage as defense mechanism against autophagic recognition. (Poster) 5th North-Regio-Day on Infection, HZI, Braunschweig, Deutschland (Oktober 2014).

Contents

Abbreviations

ABM	Actin based motility
ABP	Actin binding protein
BAP	Biotin acceptor peptide
BLI	Biolayer interferometry
CBD	Chaperone binding domain
CD	Circular dichroism
CV	Column volume
ITC	Isothermal titration calorimetry
EM	Electron microscopy
F-actin	Filamentous (polymeric) actin
G-actin	Globular (monomeric) actin
GBD	GTPase binding domain
GOI	Gene of interest
IEC	Ion exchange chromatography
LPS	Lipopolysaccharide
MS	Mass spectrometry
MST	Microscale thermophoresis
MW	Molecular weight
MWCO	Molecular weight cutoff
NLR	NOD-like receptor
NMR	Nuclear magnetic resonance
NPF	Nucleation promoting factor
PAMP	Pathogen associated molecular pattern
PCR	Polymerase chain reaction
PDB	Protein data bank
PE	Phosphatidylethanolamine
PI(3)P	Phosphatidylinositol-3-phosphate

PIP2	Phosphatidylinositol-4,5-bisphosphate
PRR	Pattern recognition receptor
RID	Rho-GTPase inactivation domain
RMSD	Root-mean-square deviation
rpm	rounds per minute
SAAT	Self-associating autotransporter
SAXS	Small angle X-ray scattering
SEC	Size exclusion chromatography
SER	Surface entropy reduction
TLR	Toll-like receptor
TTSS	Type three secretion system
VID	VirG insertion domain

Zusammenfassung

Shigella flexneri ist ein gram-negatives, humanpathogenes Bakterium, welches Shigellose bzw. Bakterienruhr verursacht. Diese fäkal-oral übertragene Krankheit tritt insbesondere in Entwicklungsländern auf und ist charakterisiert durch einen schleimigen und blutigen Durchfall. Während des Infektionszyklus leben und vermehren sich die Bakterien intrazellulär im Zytoplasma von Epithelzellen des Dünndarms. Normalerweise werden intrazelluläre Krankheitserreger vom Autophagiesystem erkannt und abgebaut. Shigellen haben jedoch einen Tarnmechanismus entwickelt, welcher ihre Erkennung und den Abbau verhindert.

Die Erkennung von Shigellen durch das Autophagiesystem basiert auf der Erkennung von VirG, ein Oberflächenprotein von *S. flexneri*, welches am alten Zellpol akkumuliert und es der Bakterienzelle ermöglicht sich durch das menschliche Zytoplasma zu bewegen. Es bildet einen Komplex mit dem menschlichen Protein N-WASP und dem Arp2/3 Komplex, welcher Aktin polymerisiert und das Bakterium vorwärts treibt. Das menschliche Protein ATG5 erkennt VirG und markiert die Bakterienzellen für den Abbau durch Autophagie. Als Gegenmaßnahme sekretiert *Shigella* das Protein IcsB, welches ebenfalls an VirG bindet und mit ATG5 um die Bindungsstelle konkurriert. Durch die höhere Affinität von IcsB, wird die Erkennung von VirG durch ATG5 verhindert, was den Tarnmechanismus von Shigellen ausmacht.

Das Ziel dieser Arbeit war die bessere Charakterisierung der molekularen Grundlagen dieses Tarnmechanismus, durch biochemische Methoden und Methoden der Strukturbiologie. Dafür sollten die Bindestellen von ATG5 und IcsB auf der Oberfläche von VirG, durch das Kristallisieren und das Lösen der Strukturen der einzelnen Proteine VirG, IcsB und ATG5, als auch ihrer Komplexe VirG/IcsB und VirG/ATG5, charakterisiert werden.

Acht initiale VirG Konstrukte lieferten ungenügende Mengen an löslichem, für die Kristallisation geeignetem Protein. Daher wurde das ESPRIT System genutzt, um nach löslichen Konstrukten der VirG α-Domäne zu suchen. Es wurden 18 interessante Konstrukte identifiziert, welche Bereiche von VirG abdeckten, für die noch keine strukturellen Informationen verfügbar waren. 13 von ihnen enthielten Teile der IcsB/ATG5 Bindedomäne. VirGE76 (340-758) war eins dieser Konstrukte, welches den Großteil der IcsB/ATG5 Bindedomäne abdeckte (außer den ersten 20 Aminosäuren). Es konnte löslich exprimiert werden und wurde in Interaktionsstudien verwendet, kristallisierte jedoch nicht. VirGE94 (419-758) war ein weiteres Konstrukt, welches die letzten 15 Aminosäuren der IcsB/ATG5 Bindedomäne beinhaltete. Es wurde kristallisiert und seine Struktur wurde durch Röntgenkristallografie gelöst.

Die VirGE94 Struktur enthielt die bereits bekannte Struktur der VirG Autochaperondomäne, eine Fortsetzung der β-Helix um etwa 4 Windungen und eine bisher unbekannte Domäne, welche hier VirG Insertionsdomäne (VID) genannt wurde, die auf der β-Helix saß. Die Faltung der VID hatte Ähnlichkeit zur Faltung der Autochaperondomänen verschiedener Autotransporterproteine. Ihre Funktion könnte daher die einer zweiten Autochaperondomäne sein, welche den Keim für eine Fortsetzung der β-Helix bildet. SAXS Experimente bestätigten, dass VirGE94 in Lösung durch eine Interaktion der VIDs dimerisierte. Diese Dimere besitzen allerdings höchstwahrscheinlich keine *in vivo* Funktion, da im längeren VirGE76 keine Dimerisierung beobachtet werden konnte und die Dimerisierungsoberfläche durch eine verlängerte β-Helix blockiert würde. Die schon früher beschriebene polare Lokalisierungsdomäne von VirG wurde als ein Teil der β-Helix identifiziert. Sie enthält eine negativ geladene Tasche, welche wichtig für die Interaktion mit einem unbekannten Partner sein könnte, welcher

für die beobachtete Verteilung von VirG verantwortlich ist. Kristallisationsexperimente von weiteren 35 VirG Konstrukten führten nicht zur Identifizierung von zusätzlichen kristallisierbaren Konstrukten, welche weiter N-terminale Bereiche der VirG α-Domäne abdeckten.

Expression und Reinigung von 37 IcsB Konstrukten ergaben in den meisten Fällen große Mengen an löslichem, homogenem Protein, allerdings konnten keine Kristallisationsbedingungen identifiziert werden. Verschiedene Strategien zur Verbesserung der Kristallisation, wie die Reduktion der Oberflächenentropie, Methylierung von Oberflächenlysinen und *in situ* Proteolyse führten ebenfalls nicht zum Erfolg. CD-spektroskopische Messungen zeigten, dass ein größerer Teil von IcsB aus Random Coil Sekundärstrukturelementen besteht, welche die Kristallisation verhindert haben könnten. Die IpgA Bindestelle von IcsB wurde im Bereich der Aminosäuren 26-125 lokalisiert, im Gegensatz zur vorher publizierten Bindestelle zwischen 171-247. Die Gegenwart von IpgA stabilisierte jene IcsB-Konstrukte, welche die IpgA Bindestelle beinhalteten, und welche in Abwesenheit von IpgA zur Aggregation neigten. Dieser stabilisierende Effekt wurde durch Thermofluor-Experimente bestätigt. Die Interaktionen von IcsB und den publizierten Interaktionspartnern Toca-1 und Cholesterin konnten hier nicht bestätigt werden. Die Ergebnisse einer Phyre2 Strukturvorhersage wurden verwendet, um eine Interaktion zwischen IcsB und N-WASP, mit einem aktivierenden Effekt auf N-WASP, vorherzusagen. Hinweise auf diese Interaktion wurden in einem Thermofluor-Assay mit N-WASP–IcsB Fusionsproteinen gefunden.

Die Kristallstruktur von humanem ATG5 konnte gelöst werden. Sie zeigte zwei Ubiquitin-artig gefaltete Domänen und eine Region mit α-helikalen Bündeln. SAXS Experimente bestätigten, dass ATG5 in Lösung monomerisch vorliegt. Die Phosphorylierung von ATG5 T75 hat einen inhibitorischen Effekt auf die Autophagie, indem die Fusion von Autophagosomen und Lysosomen verhindert wird. Hier wurden die molekularen Mechanismen dieses Effekts untersucht, indem die Strukturen einer phosphomimetischen Mutante T75E und einer Kontrollmutante T75A gelöst wurden. Es konnten jedoch keine Unterschiede in den Strukturen der ATG5 Varianten, sowie in deren Stabilitäten oder Interaktionen mit ATG16L1 festgestellt werden, was darauf schließen lässt, dass die Phosphorylierung von T75 einen eher indirekten Effekt auf die Autophagie hat, z.B. über einen unbekannten Interaktionspartner.

In Interaktionsstudien zwischen den drei am *Shigella* Tarnmechanismus beteiligten Proteinen, wurden Indizien für eine direkte Interaktion zwischen IcsB und VirGE76 durch MST gefunden. Im Gegensatz dazu konnte keine Interaktion zwischen ATG5 und einem VirG-Konstrukt beobachtet werden. Der Grund könnte in den Limitierungen der hier verwendeten Konstrukte gelegen haben, es ist jedoch auch möglich, dass das molekulare Modell des *Shigella* Tarnmechanismus modifiziert werden muss, falls zusätzliche Faktoren beteiligt sein sollten.

Summary

Shigella flexneri is a gram-negative, human-pathogenic bacterium, which causes shigellosis/bacillary dysentery. This foodborne disease occurs primarily in developing countries and manifests itself in severe bloody diarrhea. During the infection cycle the bacterial cells live and proliferate intracellularly in the cytosol of epithelial cells of the small intestine. Normally, intracellular pathogens are recognized and degraded by the autophagic system, but *S. flexneri* developed an evasion mechanism to circumvent autophagic recognition and escape degradation.

Autophagic recognition targets VirG, a surface protein of *S. flexneri*, which accumulates at one cell pole and enables the bacterial cell to move through the cytosol of human cells. It forms a complex with the human protein N-WASP and the Arp2/3 complex, which polymerizes actin and propels the bacterium forward. The human protein ATG5 recognizes VirG and targets the bacterial cell for autophagic degradation. However, as a countermeasure *S. flexneri* secretes the protein IcsB, which competes with ATG5 for binding to VirG. Due to the higher affinity of IcsB, recognition of VirG by ATG5 is blocked, thus masking *S. flexneri* from the autophagic system.

The objective of this work was a better characterization of the evasive mechanism of *S. flexneri* on the molecular level, using biochemical and structural biology methods. To this end, the binding sites of ATG5 and IcsB on the surface of VirG should be characterized, by crystallizing and solving the structures of the three individual proteins VirG, IcsB and ATG5, as well as the VirG/IcsB and VirG/ATG5 complexes.

Eight initial VirG-constructs didn't yield sufficient amounts of soluble protein, suitable for crystallization. Therefore the ESPRIT system was deployed to screen for soluble constructs of the VirG α-domain. 18 interesting constructs, covering parts of VirG, where no structural information was available yet, were identified, 13 of which contained parts of the IcsB/ATG5 binding domain. VirGE76 was one of those constructs, comprising residues 340-758, that covered most of the IcsB/ATG5 binding domain (except for the first 20 aa). It was expressed solubly and was used in interaction studies, but did not crystallize. VirGE94 was another construct, comprising residues 419-758, that covered the last 15 aa of the IcsB/ATG5 binding domain. It was crystallized and its structure was solved by X-ray crystallography.

The VirGE94 structure contained the previously published structure of the VirG autochaperone domain, a continuation of the β-helix by about four turns and a previously unknown domain, here termed the VirG insertion domain (VID), that resided on top of the β-helix. It had a fold similar to the autochaperone domains of different autotransporter proteins, which indicated that it might act as a second autochaperone domain, nucleating the fold of a continuation of the β-helix. SAXS experiments confirmed that VirGE94 formed a dimer in solution, which was mediated by an interaction of the VIDs. However, this dimer most probably does not have an *in vivo* function, as dimer formation was abolished in the longer VirGE76 construct and the dimer interface would be blocked by a prolonged β-helical backbone. The polar localization domain has been identified to be a part of the β-helical backbone of VirG and contains a negatively charged cleft that might be involved in the interaction with an unknown binding partner, which is responsible for the unipolar distribution pattern of VirG. Screening of 35 additional constructs did not result in the identification of crystallizable constructs, covering more N-terminal parts of the VirG α-domain.

Expression and purification of 37 IcsB-constructs mostly yielded high amounts of soluble and homogenous proteins, but no crystallization conditions were identified. Different rescue strategies like

surface entropy reduction, surface lysine methylation and *in situ* proteolysis were applied, but failed. CD-spectroscopy indicated that IcsB contains a major fraction of random coils that might have hindered crystallization. The IpgA binding site was mapped to IcsB residues 26-125, in contrast to the previously published binding site between residues 171-247. The presence of IpgA stabilized IcsB-constructs, containing the IpgA binding site, which were otherwise prone to aggregate. This stabilizing effect was confirmed by thermofluor-assays. The interactions of IcsB with the published binding partners Toca-1 and cholesterol could not be confirmed here. Using Phyre2 results, an interaction of IcsB and the N-WASP-GBD, with an activatory effect on N-WASP, was predicted. Indications for this interaction were observed in a thermofluor-assay, using N-WASP–IcsB fusion proteins.

The crystal structure of individual human ATG5 was solved and showed two ubiquitin-like fold domains and an α-helical bundle region. SAXS experiments confirmed that ATG5 was monomeric in solution. ATG5 T75 phosphorylation had been shown to inhibit autophagy, by the prevention of autophagosome-lysosome fusion. Here the molecular mechanisms have been investigated, by the crystallization of the phosphomimetic mutant T75E and a control mutant T75A. The structures of the ATG5 variants, their stabilities and their interactions with ATG16L1 were very similar, indicating a more indirect effect of T75 phosphorylation on autophagy, e.g. *via* a yet unidentified binding partner.

Interaction studies between the three proteins, involved in the *Shigella* camouflage mechanism, showed indications for a direct interaction between IcsB and VirGE76 by MST. In contrast, no interaction between ATG5 and any VirG-construct was observed. The reason might have been the limitations of the utilized constructs, but it might also be necessary to modify the molecular model of the *Shigella* camouflage mechanism, as additional factors might be involved.

1 Introduction

1.1 Autophagy

Autophagy is a conserved eukaryotic cellular pathway, which was first recognized in yeast as a process that recycles cellular components to sustain the metabolism in times of starvation. More recently, additional tasks of autophagy have been identified, like the degradation of aged or damaged cellular organelles and large protein complexes [1]–[3], as well as functions in the innate and adaptive immune response during infections with intracellular pathogens [4], [5]. Three different types of autophagy can be distinguished. First macroautophagy, which is characterized by the *de novo* formation of a double membrane vesicle called an autophagosome. Second microautophagy, whereby a small invagination of the lysosomal membrane is formed that segregates and degrades a small part of the cytoplasm. Third chaperone-mediated autophagy, by which proteins are transported directly into the lysosome [6]. From here on, macroautophagy is referred to as autophagy for the remainder of this work.

1.1.1 Autophagosome biogenesis

Autophagy targets cytosolic components that are engulfed by a double membrane structure, called the isolation membrane. The main source of this membrane is the mitochondria-associated endoplasmic reticulum membrane (MAM), but also other membranes contribute to it [7]. Once the isolation membrane closes around its cargo, the vesicle is called an autophagosome. Subsequently the outer membrane of the autophagosome fuses with a lysosome, forming an autolysosome, and the lysosomal enzymes degrade the inner membrane and the autophagosomal cargo [6]. Degradation products like amino acids are transported across the autolysosomal membrane and are recycled [8] (Figure 1).

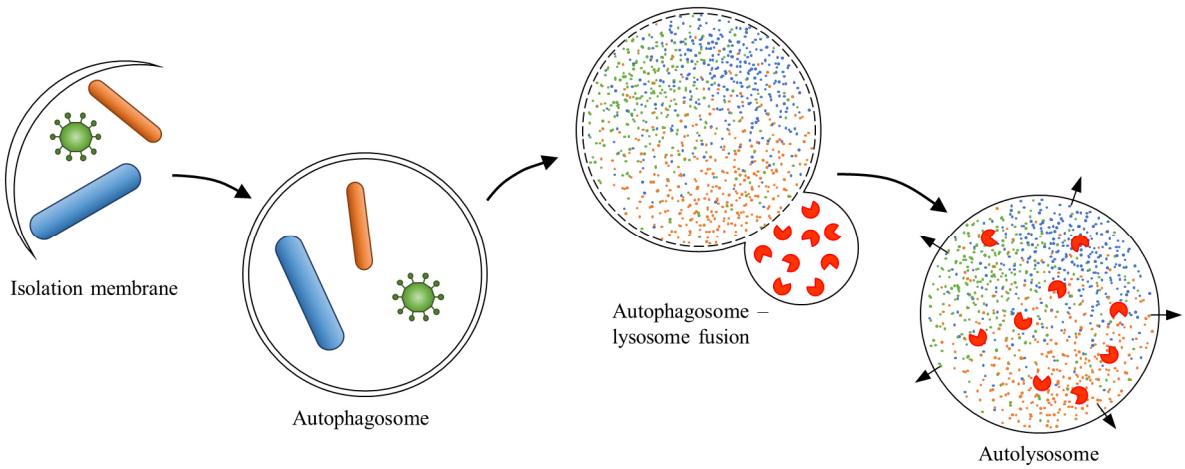


Figure 1: Scheme of the autophagic pathway. Cytosolic components are engulfed by an isolation membrane. The closed vesicle is called an autophagosome, which then fuses with a lysosome. The contents of the formed autolysosome are degraded and the degradation products like amino acids are transported across the membrane to be recycled (modified from [9]).

In yeast, more than 30 genes have been identified that are involved in the autophagy pathway [10]. TOR is the key regulator of starvation induced autophagy. It is a serine/threonine kinase, which is active under nutrient rich conditions and inactive under starvation conditions. It regulates the Atg1 complex, which plays an important role in the initiation of autophagy. Atg1 is a serine/threonine kinase whose substrates are still unknown. It forms a complex with Atg13, Atg17, Atg29 and Atg31. Atg13 is hyper-phosphorylated by the active TOR kinase, which leads to the inhibition of Atg1 kinase activity and consequently to the downregulation of autophagy under nutrient rich conditions [11].

A second protein complex, comprising Vps34 (a phosphatidylinositol kinase), Vps15, Vps30/Atg6, Atg14 and Atg38, is responsible for the generation of phosphatidylinositiol-3-phosphate (PI(3)P) in the autophagosomal membrane [12], [13]. PI(3)P is a phospholipid that recruits effector proteins to the

autophagosomal membrane. One example is the Atg5-12/16 complex, which plays a role in the conjugation of Atg8 to phosphatidylethanolamine (PE). The formation of this complex and the conjugation of Atg8 to PE are described in the following section.

Two ubiquitin-like conjugation systems, which are essential for autophagy [14], play a role in the expansion of the isolation membrane [15]. The first system conjugates the ubiquitin-like protein Atg12 to Atg5 *via* an isopeptide bond, involving the C-terminal G186 of Atg12 and the K149 sidechain of Atg5 [16]. Atg7 acts as the E1 activating enzyme [17] and Atg10 as the E2 conjugating enzyme [18], but no typical E3 ligase is involved. The Atg5-12 conjugate interacts with Atg16. Atg16 dimerizes and a complex involving two copies of Atg5-12/16 is formed [19], [20]. The second system conjugates the ubiquitin-like protein Atg8 to PE [21]. Atg7 also acts as the E1 activating enzyme in this second process, while the E2 conjugating enzyme is Atg3 [21]. The Atg5-12/16 complex, formed by the first conjugation system, functions as the E3 ligase in the second process [22] (Figure 2). Orthologues of each component of the two ligation systems are also present in human cells [23]. For instance ATG16L (two isoforms L1/L2) is the human orthologue of yeast Atg16 [24] and LC3A (two isoforms)/B/B2/C, as well as GABARAP/L1/L2 are the seven human orthologues of yeast Atg8 [25].

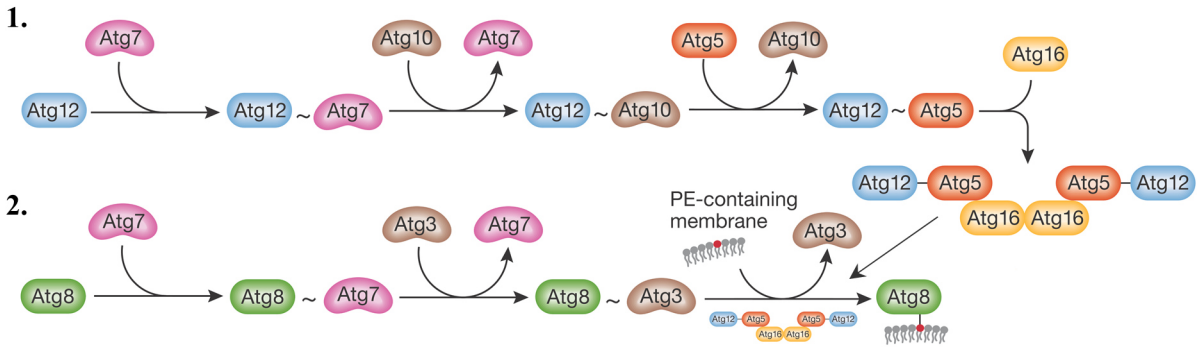


Figure 2: The two ubiquitin-like conjugation systems in autophagy. In the first system, Atg12 is conjugated to Atg5 by Atg7 and Atg10. The second system conjugates Atg8 to PE through the actions of Atg7, Atg3 and the Atg5-12/16 complex (modified from [26]).

The Atg5-12/16 complex interacts with PI(3)P [20] and localizes to the outer surface of the growing isolation membrane [27]. As it promotes the lipidation of Atg8, it has a function in targeting the Atg8-PE conjugate to the correct membrane [28]. After the completion of the autophagosome it rapidly dissociates from the membrane [29]. Atg8-PE localizes to both sides of the growing isolation membrane. It mediates membrane hemifusion, which is important for the expansion of the growing isolation membrane [30], and it probably functions as a scaffolding protein that supports the growing membrane [26]. Like the Atg5-12/16 complex it dissociates from the outer surface of the completed autophagosome [31], but it remains attached to its inner membrane [32] (Figure 3).

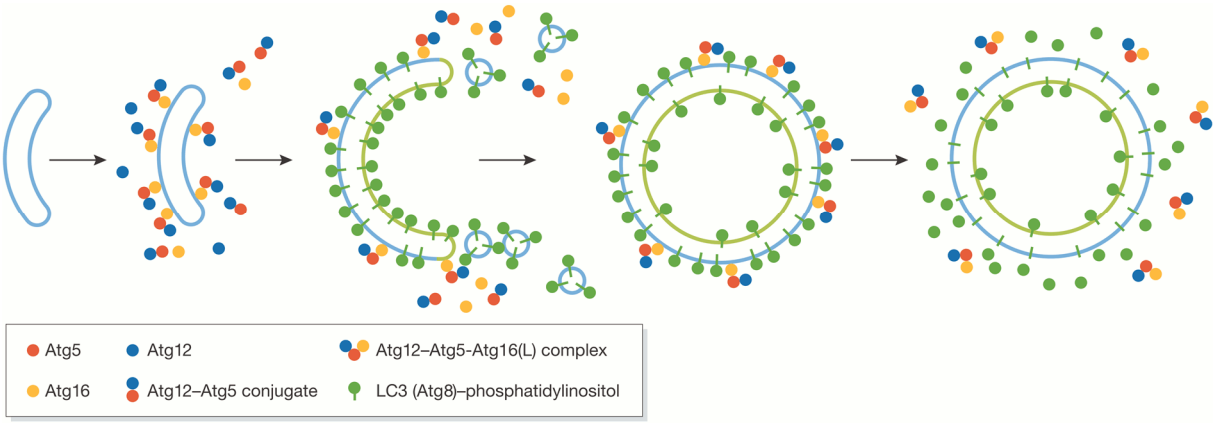


Figure 3: The Atg5-12/16 complex is associated with the outer surface of the growing isolation membrane, while Atg8-PE localizes to both sides. Upon closure of the vesicle, the Atg5-12/16 complex and Atg8 dissociate from the outer surface of the autophagosome, but Atg8 stays associated with the inner membrane [26].

1.1.2 Autophagy in infection and disease

Intracellular pathogens have to be detected by the cell, before they are degraded by autophagy. Pattern recognition receptors (PRRs) detect compounds that are characteristic for pathogens, called pathogen associated molecular patterns (PAMPs). PRRs comprise different proteins like the membrane-bound Toll-like receptors (TLRs) and the cytosolic NOD-like receptors (NLRs). For instance, lipopolysaccharides are detected by TLR4, while ssRNA is recognized by TLR7 [33], [34]. In both cases autophagy is activated in subsequent steps [4]. Nod1 and Nod2, which recognize the bacterial peptidoglycan constituents D-glutamyl-meso-diaminopimelic acid and muramyl dipeptide, recruit Atg16L1 to the plasma membrane at the entry site of *S. flexneri*, thus targeting autophagy to invading bacterial cells [35]. Beside the stimulation of autophagy, PRRs also activate either the NF-κB innate immune response for the production of proinflammatory cytokines or caspase 1 for the induction of programmed cell death [36].

Following their recognition, membrane structures surrounding the bacteria, and possibly also the bacterial surface itself, are polyubiquitinated [37], which is a general degradation signal. Still, the responsible molecular systems, as well as the protein targets for ubiquitination remain elusive. Only recently, the E3 ubiquitin-ligases parkin and LRSAM1 have been identified, which play roles in the polyubiquitination of *Mycobacterium tuberculosis* and *S. flexneri* [38], [39]. Different adaptor proteins like p62, NBR1, NDP52 and OPTN direct the growth of the isolation membrane around the bacterium, by binding to both the polyubiquitin signals, associated with the bacteria, and the LC3 on the inner surface of the growing isolation membrane [40]. Upon closure of the vesicle, the autophagosome fuses with a lysosome and the bacterium is degraded by the lysosomal enzymes, as described above. A different pathway targets bacteria, which are about to escape the phagosome. Galectin 8 binds to both host glycans, which are exposed on damaged *Shigella* containing vacuoles, and to NDP52. This way antibacterial autophagy is targeted to the escaping bacterium [41].

1.1.3 Bacterial autophagy evasion strategies

Despite the described antimicrobial pathways, some intracellular pathogens circumvent degradation by autophagy and others even exploit the autophagic mechanisms for their own benefit. For instance, *Anaplasma phagocytophium*, *Coxiella burnetii*, *Legionella pneumophila*, *Mycobacterium spp.*, *Staphylococcus aureus* and *Yersinia pestis* first induce their uptake by an autophagosome, but then inhibit autophagosome-lysosome fusion. This way they create a vacuolar replicative niche [42]. Other bacteria developed mechanisms for the down regulation of autophagy. The ubiquitination of infection-specific aggresomes usually leads to a general increase in autophagy. *Salmonella spp.*, *L. pneumophila* and pathogenic *Escherichia coli* dampen the autophagic response, by blocking the ubiquitination of

aggresomes or by expressing specific deubiquitinases [42]–[44]. Another example is *L. pneumophila*, which secrete the RavZ protein that inhibits autophagy by the deconjugation of LC3 from PE. RavZ cleaves LC3 one residue N-terminal of the conjugation point, which also prevents reconjugation [45]. Finally, *S. flexneri* and *Listeria monocytogenes* developed camouflage mechanisms to avoid recognition by autophagy, by means of secreted proteins (IcsB, InlK) [46], [47] (Figure 4). The exact mechanism by which *S. flexneri* avoids autophagic recognition is described in greater detail in chapter 1.3.1.

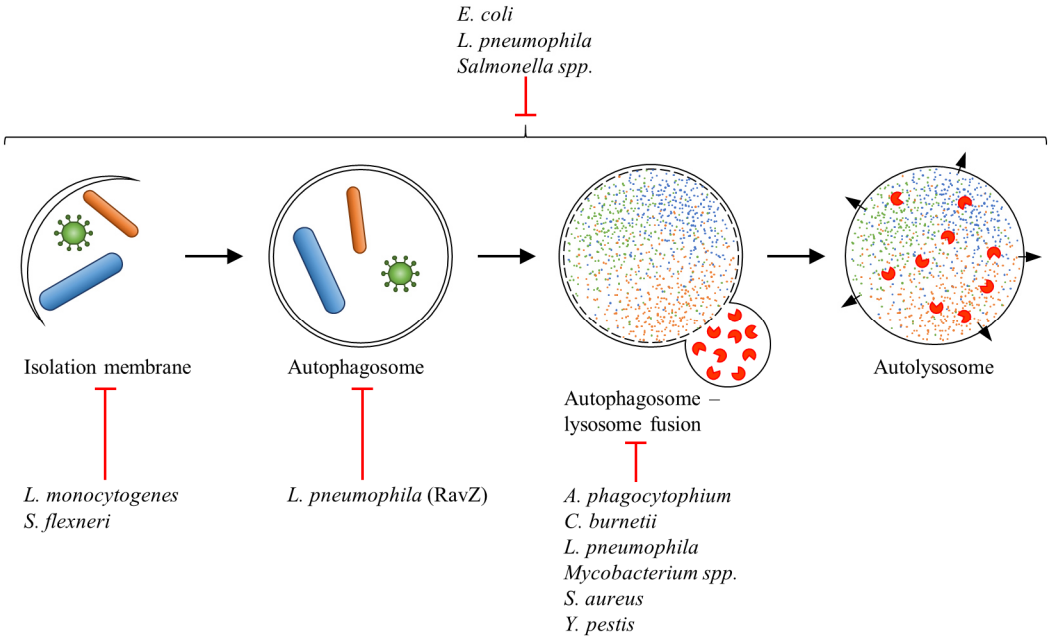


Figure 4: Intracellular pathogens avoid autophagic degradation by different means. *L. monocytogenes* and *S. flexneri* developed a camouflage mechanism to avoid autophagic recognition. *L. pneumophila* RavZ deconjugates LC3 from PE and inhibits the maturation of the autophagosome. *A. phagocytophium*, *C. burnetii*, *L. pneumophila*, *Mycobacterium spp.*, *S. aureus* and *Y. pestis* induce their uptake by an autophagosome, but inhibit autophagosome-lysosome fusion at a later stage. Finally pathogenic *E. coli*, *L. pneumophila* and *Salmonella spp.* downregulate the autophagy system, by avoiding aggresome ubiquitination (reproduced from [9]).

1.2 The actin cytoskeleton

A number of pathogens have adopted to an intracellular live style by developing an actin based motility (ABM), thus exploiting the eukaryotic actin polymerization system that is described in this chapter. Beside the intermediate filaments, the microtubules and the septin filaments, the actin- or microfilaments are a component of the cytoskeleton with important functions in cellular motility and as tracks for myosin motor proteins. The actin filament network can generate forces, when the growing ends of actin filaments push against barriers like cellular membranes, which is necessary for cellular motility. Actin naturally occurs in two different states, as monomeric globular (G) actin and as polymeric filamentous (F) actin [48].

1.2.1 G-actin

G-actin is a highly conserved 375 amino acid protein with a MW of about 42 kDa. Vertebrates express six different actin isoforms: four α-isoforms (expressed in skeletal, cardiac and smooth muscle tissue), one β-isoform and one γ-isoforms (mainly expressed in non-muscle tissue) [48]. The first G-actin structure has been solved in 1990 in complex with DNase I [49]. Actin is composed of four different subdomains. Subdomains 1 and 3 are similar in structure, while subdomains 2 and 4 can be viewed as large insertions into the former two subdomains. Between subdomains 1/2 and 3/4 the hinge region, the nucleotide cleft and the target binding cleft are located [50]. The nucleotide cleft binds a nucleotide (either ADP or ATP) and a divalent cation (Mg^{2+}). The hydrophobic target binding cleft is the binding site for most actin binding proteins (ABPs) [51]. Subdomain 2 carries the DNase I-binding (D) loop that

plays an important role for the inter-subunit contacts in the filament, as it inserts into the target binding cleft of the neighboring subunit [50] (Figure 5).

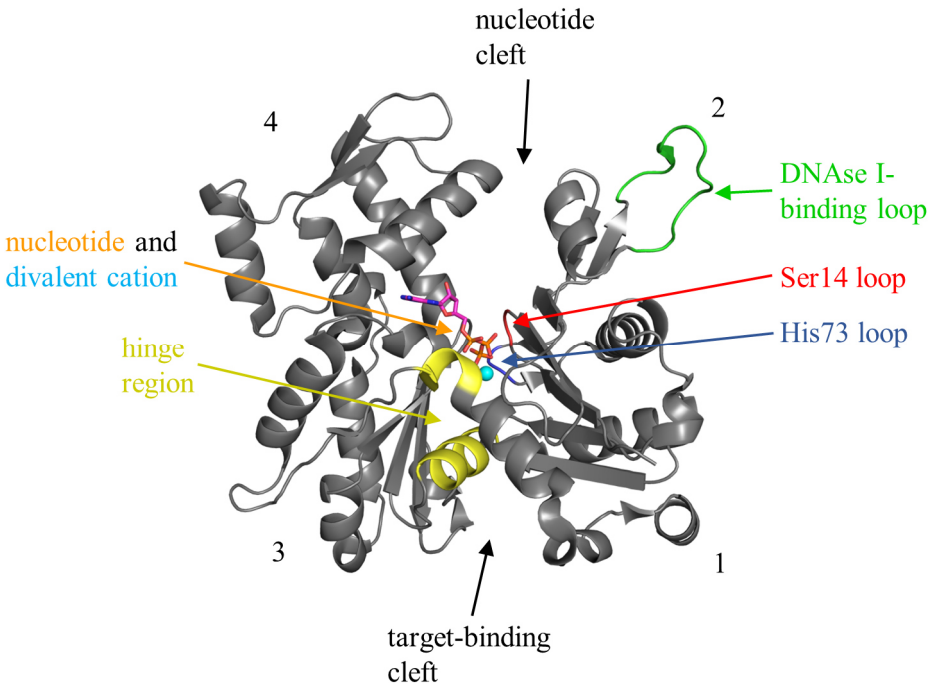


Figure 5: G-actin structure from PDB-entry 2A42. The numbers designate the four subdomains. A detailed description of the structural features is given in the text.

Only minor conformational changes can be observed between the ADP- and ATP-bound form of G-actin, mainly involving the Ser14 β-hairpin loop and the sensor loop, carrying the methylated His73. Both loops contact the phosphates of the nucleotide. The conformational changes of the two loops propagate to subdomain 2 with the D-loop, which also shows only minor conformational changes. Still these differences in the structure seem to be the major cause for a decreased stability of F-actin in the ADP-bound state [52]. The actin monomers adopt a flatter conformation in F-actin than in G-actin, which is due to a rotation of subdomains 1/2 by about 20° with respect to subdomains 3/4 [53].

1.2.2 F-actin

F-actin is composed of helically assembled actin monomers, which form the filament with a diameter of about 7 nm [48]. Current atomic models of the actin filament are derived from X-ray fiber diffraction of oriented actin gels or from cryo electron microscopy (EM). The resolution of initial X-ray fiber diffraction derived structures was only about 7 Å [54]. But actin fibers are diamagnetic and when they were placed in a strong magnetic field their order was increased and a resolution of 3.3 Å in radial and 5.6 Å in equatorial direction was obtained [53]. The best cryo EM derived structure reaches a resolution of 6.6 Å [55]. These structures show that F-actin forms a left-handed helix of actin monomers with a repetitive motif, consisting of about 13 molecules that form six consecutive turns and cover an axial distance of 359 Å. The twist per molecule is -166.6°, which is close to -180°. Consequently the structure appears like two slowly turning right-handed chains (Figure 6).

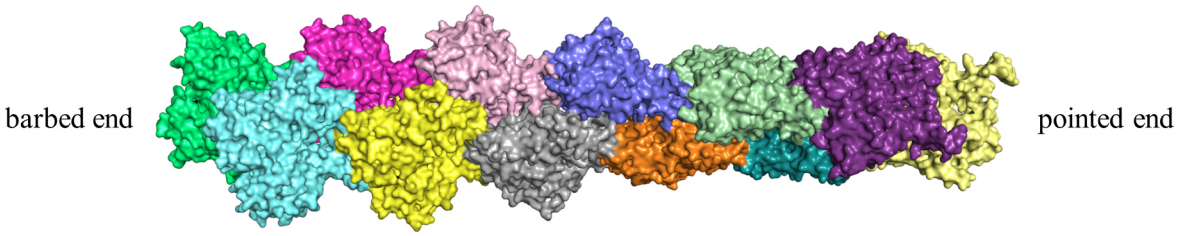


Figure 6: Structure of F-actin from PDB-entry 3G37. Actin monomers form a left handed helix with a twist of -166.6° per molecule. As this value is close to -180°, the filament appears like two intertwining right handed helices.

Actin has an ATPase function, but ATP hydrolysis is not very efficient in G-actin compared to F-actin, which readily hydrolyzes ATP to ADP with a half-life of 2 s [50], [56]. Subsequently, the γ-phosphate slowly dissociates with a half-life of 6 min [57]. F-actin is a polar filament and has two inequivalent ends, a barbed/+ end and a pointed/– end (Figure 6). The two ends have different critical actin concentrations, which indicate the concentration of G-actin, where the polymerization rate equals the depolymerization rate. The critical concentration of the barbed end is 0.1 µM, while the critical concentration of the pointed end is 0.6 µM. Consequently actin monomers usually join the barbed end of a filament in the ATP-bound state with diffusion limited rate, then hydrolysis takes place in the filament and finally ADP-bound actin dissociates from the pointed end [58]. In solution, ADP is replaced by ATP and actin can reassociate with the barbed end. This process is called actin filament tread milling [59] (compare Figure 7).

1.2.3 Actin binding proteins (ABPs)

In vivo, actin is present at high intracellular concentrations of 20-100 µM, which lies far above the critical concentrations. Therefore actin filament formation needs to be tightly regulated, which is accomplished by the action of different ABPs. The dynamics of the actin cytoskeleton are controlled by small Rho-GTPases. Rho itself controls the formation of stress fibers, Rac the formation of lamellipodia and Cdc42 the formation of filopodia. They regulate a number of ABPs, which are responsible for the nucleation, elongation, stability, depolymerization and different other aspects of the actin cytoskeleton. For instance, Tβ4 and profilin are actin sequestering proteins. They bind monomeric G-actin and prevent its uncatalyzed nucleation. Profilin promotes the exchange of ADP for ATP. It binds to the barbed end of actin monomers, thereby leaving the pointed end free for barbed end growth of the filament, while inhibiting pointed end growth. Besides binding to actin monomers, profilin also binds proline rich sequences, which are present in several ABPs, e.g. the Arp2/3 complex and formins (see below). It thus delivers actin monomers for the formation of filaments. Another ABP is the actin depolymerization factor cofilin, which severs and depolymerizes actin filaments in the ADP-bound state. Different capping proteins like CapZ or tropomodulin bind to the barbed or pointed ends of the filament and stabilize them. Other proteins like filamin and fascin act as actin filament cross linkers. Finally, a number of proteins and complexes are able to initiate actin filaments, which include the ubiquitously expressed Arp2/3 complex and formins. The expression of other initiating proteins like Spire, Cobl, VopL/VopF, TARP and Lmod is limited to special tissues [48].

1.2.3.1 The Arp2/3 complex

The Arp2/3 complex initiates a filament branch, starting from an existing filament. The resulting branched filament network pushes forward the leading edge of motile cells (Figure 7).

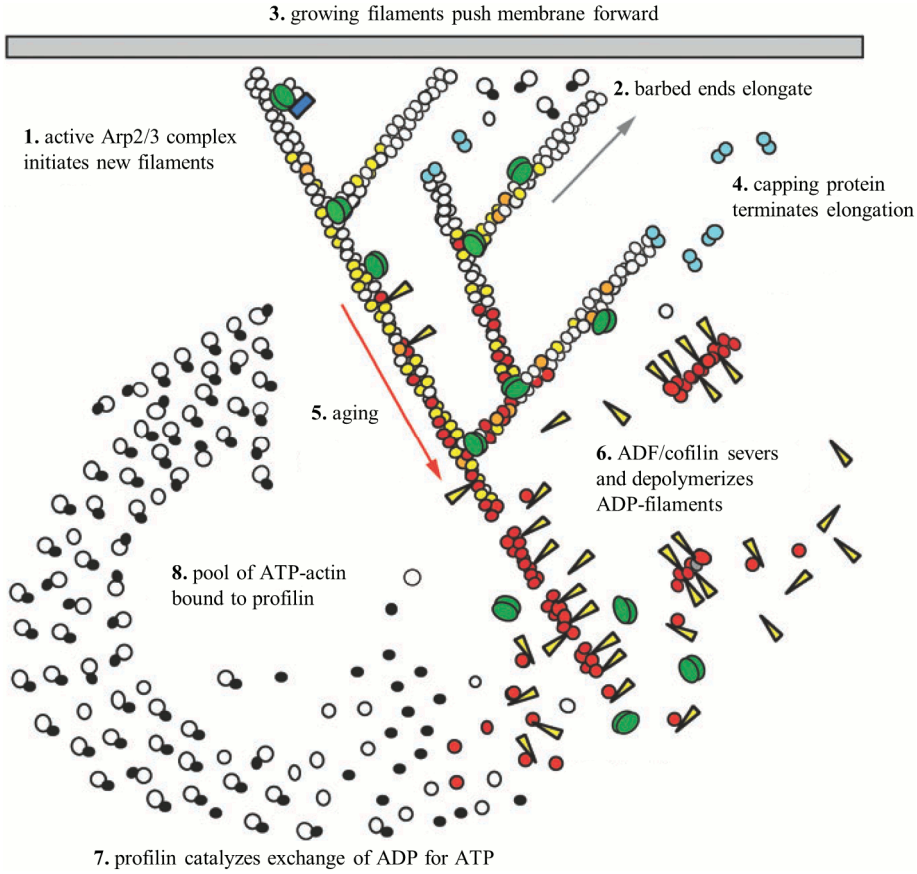


Figure 7: 1. The active Arp2/3 complex initiates new actin filaments, starting from an existing filament. 2./3. The growing barbed ends of the filaments push against the plasma membrane, forming a lamellipodium. 4. The barbed end growth of some filaments is stopped by capping proteins. 5. When the actin monomers join the barbed ends of the filaments, they are ATP bound. ATP is quickly hydrolyzed to ADP + P_i and the phosphate dissociates from the filament. 6. ADF/cofilin severs and depolymerizes ADP-bound filaments. 7. The emerging ADP-bound actin monomers bind to profilin, which catalyzes the exchange of ADP for ATP. 8. Finally ATP-bound actin-profilin complexes rejoin the barbed end. This circulation of actin is called actin filament treadmilling (adapted from [60]–[62]).

The Arp2/3 complex consists of seven subunits: Arp2, Arp3 (actin related subunits), p16, p20, p21, p34 and p40. The two actin related subunits are stabilized in an inactive state by the five other subunits. The activation of the complex is accomplished by the cooperative action of nucleation promoting factors (NPFs), an actin filament and an actin monomer. The structure of the inactive complex has been solved in 2001 (Figure 8 a), but up to date there is no structure of the active complex. Arp2 and Arp3 structurally resemble actin but the nucleotide cleft is more open and lacks a bound nucleotide. The Arps lie head to tail, like subunits in an actin filament, but they are rotated by -180° around the filament axis relative to each other (not -166.6° like in a filament). In this conformation the nucleation of an actin filament is not possible. A proposed model for the conversion of the inactive into an active complex includes 1) the binding of two nucleotides to Arp2 and Arp3 to close their nucleotide clefts and 2) a rotation of about 20°, which rearranges the Arps into a conformation that acts as nucleation template of a new filament (Figure 8 c). The Arps are then the first subunits in the new branch (Figure 8 b). The role of the NPFs might be the stabilization of the active complex by bridging p21, p40 and the pointed ends of the Arps. After the activation of the Arp2/3 complex, the NPFs dissociate in a sub-second timescale [62] [63].

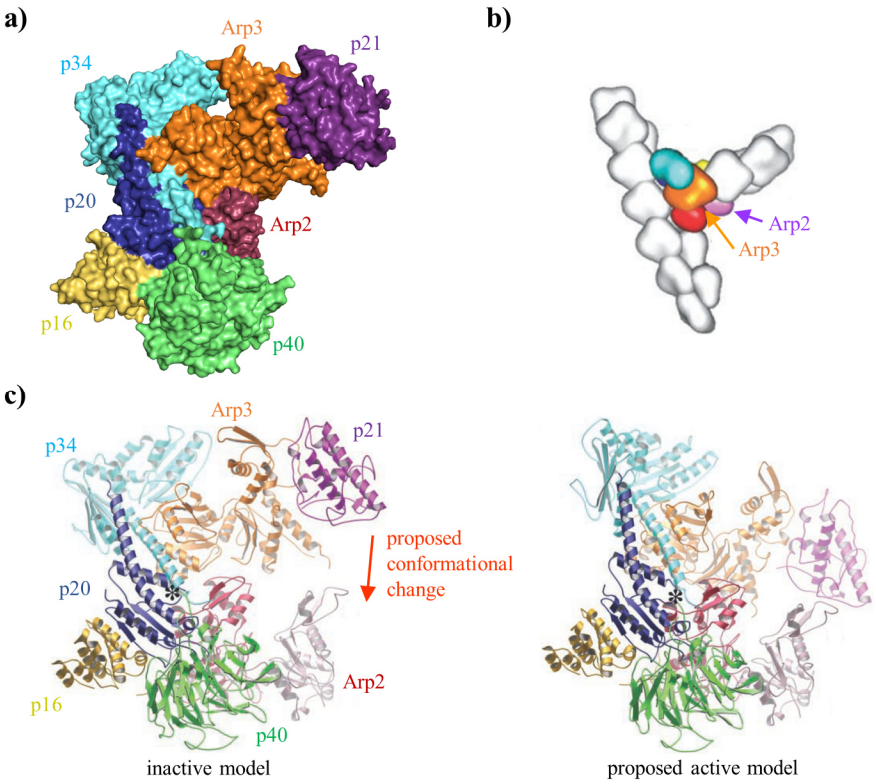


Figure 8: a) Structure of the Arp2/3 complex from PDB-entry 1K8K. b) Reconstruction of an actin filament branch, including the Arp2/3 complex, from electron micrographs [62], [64]. c) Proposed model for the activation of the Arp2/3 complex [62], [63].

1.2.3.2 Formins and other initiating proteins

Formins nucleate and elongate unbranched filaments for actin bundles in filopodia and the contractile ring. In contrast to the Arp2/3 complex, a previously existing filament is not necessary. In humans, 15 different formins exist, which are grouped into eight families [65]. Formins form homodimers. They contain several domains including formin homology 1 (FH1), FH2 and regulatory domains. FH1 domains contain one or more short proline rich sequences, which bind profilin-actin complexes and accumulate them near the end of the growing filament. Two FH2 domains form a ring that is wide enough to accommodate two actin subunits (Figure 9 a). They can nucleate actin polymerization, by stabilizing actin dimers and stay associated with the barbed end of the growing filament over thousands of subunits and seconds, a property called processive capping. Other domains are more diverse. Diaphanous related formins contain an N-terminal diaphanous inhibitory domain (DID), also called GTPase binding domain (GBD), that binds to a C-terminal diaphanous autoregulatory domain (DAD), resulting in autoinhibition similar to N-WASP. Activated Rho can bind to the GBD, which releases the DAD and leads to the activation of the formin [66]. A current model of the elongation mechanism of formins claims that they adopt two different states, an open and a closed one. After the addition of each actin subunit, the formin transiently adopts the closed or capped state to maintain its position at the end of the filament. Profilin increases the elongation rate, depending on the number of profilin-binding sites in FH1. The elongation rate can even increase over the diffusion limited rate of free barbed ends. The profilin binding sites in FH1 create higher local actin concentrations, supporting this elevated rate (Figure 9 b) [62].

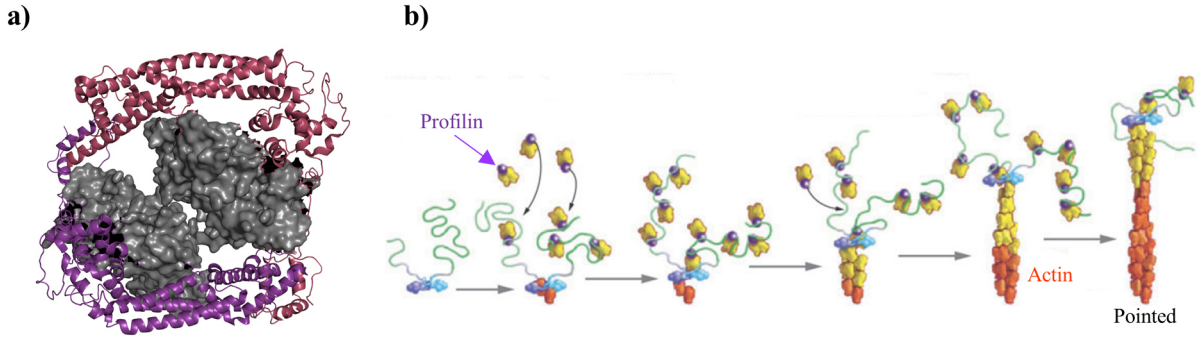


Figure 9: a) Structure of a FH2-dimer (red, violet) in complex with two actins (grey) from PDB-entry 1Y64. b) Processive barbed end growth of an actin filament, catalyzed by a formin [67].

Finally Spire, Cobl, VopL/VopF, TARP and Lmod nucleate actin filaments by means of their tandem WH2-domains. One WH2 domain consists of 17-27 amino acids that form an N-terminal helix, which binds the target-binding cleft of an actin monomer, and a C-terminal linker. By the binding of three to four actin subunits the local actin concentration is increased and a nucleus is formed more readily. These proteins only catalyze the nucleation, not the elongation of an actin filament [68].

1.2.3.3 Nucleation promoting factors (NPFs) and Toca-1

NPFs include proteins of the Wiskott-Aldrich syndrome protein (WASP) family like WASP, N-WASP and Scar/WAVE [69]. WASP and N-WASP have most of their domains in common: a pleckstrin homology (PH) domain that binds phosphatidylinositol-4,5-bisphosphate (PIP2), an IQ motif that binds calmodulin, a GBD that binds Cdc42, a proline-rich region (PRR) and a VCA domain including verprolin homology/WASP homology 2 (V/WH2), connecting/cofilin homology (C) and acidic (A) subdomains [70]. The VCA domain has been shown to activate the Arp2/3 complex [71]. V and C have an affinity for actin, while C and A interact with the Arp2/3 complex. The acidic A-domain probably binds to basic surface patches on p21 and p40 and stabilizes the Arp2/3 complex in an active conformation, while at the same time the first actin monomer of the new filament is anchored by the V-domain in the correct position. This presents an effective strategy, as actin trimers are normally very unstable [63]. In an unbound state N-WASP is complexed with N-WASP interacting protein (WIP) and stabilized in an autoinhibited conformation [72], [73], where an intramolecular interaction between the GBD and the C-domain inhibits the activity of the VCA-domain. N-WASP can be activated by the cooperative action of Cdc42 and Toca-1. Upon interaction, the VCA domain is released and is free to interact with and thereby activate the Arp2/3 complex [74].

Toca-1 is a protein that together with CIP4 and Fbp17 belongs to the Pombe Cdc15 homology (PCH) family. It contains an F-BAR domain, an HR1 domain and a SH3 domain. F-BAR domain proteins usually form dimers and play important roles in the remodeling of membranes [75]. The HR1 domain adopts a coiled coil fold and has been shown to bind Cdc42 [74]. Finally the SH3 domain interacts with proline rich sequences in both N-WASP and formins [74], [76], [77]. Hence, Toca-1 is involved in the regulation of actin polymerization by both the Arp2/3 complex and formins [78].

1.2.4 Actin based motility (ABM)

Several intracellular pathogens exploit the actin polymerization system to move through the cytosol, a process called actin based motility. A number of different mechanisms have evolved for this purpose. *L. monocytogenes* is the causative agent of listeriosis, which manifests as gastroenteritis or as infection of the central nervous system. *Listeria* are capable of ABM by means of their outer membrane protein ActA, which directly activates the Arp2/3 complex by mimicking the action of activated N-WASP. This leads to the formation of actin tails and the generation of a propulsive force, pushing the bacterial cells through the cytoplasm [79]–[81]. Different pathogenic *Rickettsia* species, the causative agent of spotted

fever, express the outer membrane protein Sca2, which is a bacterial mimic of formins. In contrast to the branched tails of *Listeria*, they are generating propulsion by the formation of linear actin tails [82], [83]. Also the vaccinia virus is capable of ABM *via* its surface protein A36R that interacts with an N-WASP accessory protein like Nck or WIP. Consequently, N-WASP and the Arp2/3 complex are activated [84], [85].

Finally, also *Shigella* species are capable of ABM [86], which in this case is mediated by the outer membrane protein VirG/IcsA [87], [88]. VirG localizes to the old cell pole where it interacts with and activates N-WASP, which in turn activates the Arp2/3 complex [89]. A branched actin tail, similar to the one of *Listeria*, is formed that pushes the bacterial cell through the cytoplasm (Figure 10). When a *Shigella* cell reaches the plasma membrane, a protrusion is formed, reaching into the neighboring cell. The protrusion is taken up and *Shigella* can escape into the cytoplasm of a new cell. Efficient intracellular motility requires N-WASP and Arp2/3, but efficient intercellular spread requires formins. Actin tails in the cytoplasm consist of branched filaments, indicating a formation by the Arp2/3 complex, while actin tails in protrusions consist of parallel filaments, indicating a formation by formins. Indeed, the inhibition of the formin Dia1 prevents the formation of protrusions and inhibits intercellular spread, but not intracellular motility. Dia1 colocalizes with *Shigella* cells in protrusions, but not in the cytoplasm. IpgB2, a protein secreted by *Shigella*, has been shown to be able to bind and activate Dia, but deletion mutants are not deficient in intercellular spread, indicating the existence of other means for the activation of formins [90]. Consequently, for a successful infection *Shigella* has to manipulate both the Arp2/3 and the formin actin polymerization systems. The next chapter introduces *Shigella* in more detail and describes the infection mechanism.

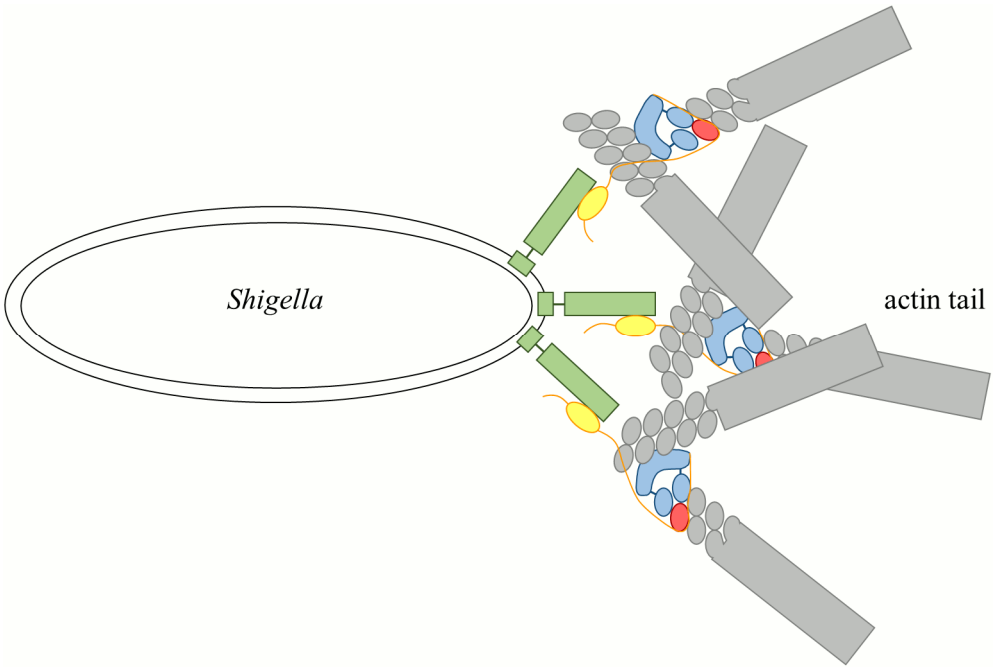


Figure 10: The *Shigella* outer membrane protein VirG (green) activates N-WASP in a manner similar to Cdc42, by binding to its GBD (yellow circle). The N-WASP VCA domain (orange line) in turn activates the Arp2/3 complex (blue), by stabilizing a filament nucleus, comprising Arp2, Arp3 (blue circles) and an actin monomer (red circle). Starting from these nuclei, actin filaments (grey) are forming a branched actin tail.

1.3 Shigella

The genus *Shigella* comprises four different species: *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei*. All of them can cause shigellosis, also called bacillary dysentery, which is a severe bloody diarrhea. In Germany, 552 cases were reported to the Robert Koch Institut in 2014 [91], while in the United States, 7746 cases were reported to the Centers for Disease Control and Prevention in 2012 [92]. In both countries, the main cause of infection was *S. sonnei*. Taken these numbers, shigellosis is not a major health problem in industrialized countries, but it does pose a major problem in developing countries. Shigellosis is endemic in most developing countries [93] with 160 million cases and 1.1 million deaths per year worldwide, mainly caused by *S. flexneri*, as estimated in 1999 [94]. The major mode of transmission is the fecal-oral route [95]. Hence, the main problems are poor sanitation and hygienic standards, the lack of clean drinking water, malnutrition and the costs of antibiotic treatment.

Shigella are gram-negative, facultative anaerobe, non-motile and non-encapsulated bacteria. In the absence of flagellar or capsular antigens, the antigenic characterization is based on the O-antigens. Accordingly, *S. dysenteriae* (group A) are subdivided into 15 sero-types, *S. flexneri* (group B) into 15 sero- and sub-types, and *S. boydii* (group C) into 20 sero-types [96]. *S. sonnei* (group D) comprises just 1 sero-type, subdivided into 5 bio-types [97]. Considering that the major cause of *Shigella* infections is *S. flexneri* and the dominant sero-type is 2a, most research on *Shigella* infections has been performed on this sero-type and additionally on sero-type 5a. *Shigella* and *E. coli* are highly similar, both biochemically and on the genetic level. Accordingly they can be considered as members of the same species in the genus *Escherichia*. The current classification is kept for historical reasons [98].

The host immune response is sero-type specific and results in immunity against infections with the same sero-type. Thus vaccination is possible in principle [95]. Currently, different approaches to develop vaccines are being followed. These include the development of vaccines directed against the O-antigen, conjugate vaccines and conserved-protein-directed vaccines. Still, up to date no vaccine is available being effective against all different sero-types [96]. Currently, *Shigella* infections are treated with antibiotics, whose continuous use is compromised by an increasing number of resistant strains [99]. The lack of a vaccine and the increasing number of resistant strains shows the importance of the identification of new ways to control *Shigella* infections.

All pathogenic strains share a 220 kb virulence plasmid, that carries most of the virulence factors [100] including all necessary components of a type three secretion system (TTSS), translocators (IpaB, IpaC), transcription activators (VirF, VirE, MxiE), effector proteins (IcsB, IpaA, IpaB, IpaC, IpgB1, IpgB2, IpgD, VirA) and chaperones (IpgA, IpgC, IpgE, Spa15) [101]. The expression of the virulence genes is controlled by the environmental temperature. Expression is upregulated upon a temperature shift from 30°C to 37°C, which corresponds to an uptake of the bacterial cell into the human body [102]. Responsible for the temperature regulation is the histone-like protein 1 (H-NS protein) [103]. The bacterial TTSS is a complex machine, which is able to transport a set of about 25-30 effector proteins from the bacterial cytoplasm directly into the host cytoplasm [104]. It is composed of a cytoplasmic part, a transmembrane domain, spanning the inner membrane, the periplasm with the peptidoglycan layer and the outer membrane, and an extracellular needle [105]. It is activated upon contact to the host cell membrane.

A number as small as ten *Shigella* cells can cause an infection in humans [106]. The bacterial cells pass the acidic milieu of the stomach by an unknown mechanism to reach the small intestine. The presence of the bile salt deoxycholate (DOC) is a stimulus for *Shigella* to prepare for infection. IpaB is recruited to the tip of the TTSS needle by a conformational change of IpaD, induced by the binding of DOC [107]. Also, in the presence of DOC, the *Shigella* surface protein VirG acts as an adhesin, that mediates the contact of bacterial and human cells *via* an unknown receptor [108]. *Shigella* cells have three known

ways to pass the epithelial barrier: 1) by translocation through M-cells [109], 2) by passing between the epithelial cells after manipulation of the tight junctions and 3) by passing through gaps created by polymorphonuclear leucocytes in response to interleukins [95]. In the solitary lymphatic nodules, underlying the M-cells, they infect macrophages [110], which undergo induced apoptosis [111]. The released *Shigella* cells then infect enterocytes *via* the basolateral membrane. They induce membrane ruffling to be taken up by phagocytosis. The internalized cells immediately disrupt the phagocytic membrane and are released into the cytoplasm, where they are motile by ABM and proliferate. When bacterial cells reach the plasma membrane, protrusions are formed reaching into the neighboring epithelial cells. These cells take up the bacteria in vacuoles, which again are immediately disrupted. This way, *Shigella* disseminates from cell to cell without being exposed to the environment, thus circumventing the humoral immune response (Figure 11).

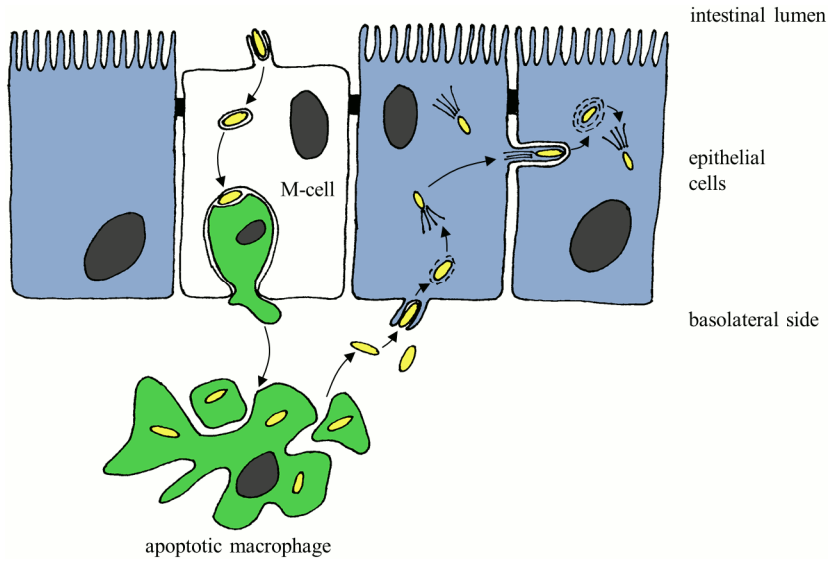


Figure 11: *Shigella* cells translocate through M-cells and infect macrophages, where they proliferate and later induce apoptosis. The released bacterial cells induce their phagocytic uptake by epithelial cells *via* the basolateral membrane. Once inside they immediately disrupt the phagosome and are motile by ABM. When a plasma membrane is touched, a protrusion reaching into the neighboring cell is formed, which is taken up by this cell. *Shigella* escapes into the cytoplasm, thus infecting a new cell and spreading the infection [112].

Shigella not only evades the humoral immune response, additionally it uses a number of ways to modulate the innate immune response. *Shigella* lipopolysaccharides (LPS) and peptidoglycans are the source of a strong inflammatory response of the innate immune system. The inflammation causes the epithelial barrier to break down, resulting in an increased invasion of bacterial cells from the lumen of the small intestine [113]. A number of effector proteins secreted by the TTSS are also involved in the modification of the innate immune response. OspG is a kinase that binds to a set of ubiquitin conjugating enzymes, especially those involved in the degradation of I-κB, an inhibitor of NF-κB [114]. OspF, a phosphothreonine lyase, inactivates the MAP kinases p38 and Erk1/2. This results in the inhibition of a set of NF-κB controlled genes, leading to the repression of the innate immune response [115]. $IpaH_{9.8}$ has an E3 ubiquitin ligase function and binds the splicing factor $U2AF^{35}$, resulting in a modulation of the inflammatory response [116], [117]. All these factors act together in the modulation of the innate immune response to allow for the most efficient propagation of the *Shigella* infection in the tissue.

A number of other effectors play important roles in the uptake of the bacterial cell by phagocytosis, including IpaA, IpaB, IpaC, IpgB1, IpgB2, IpgD and VirA [118]. IpaA is an effector with multiple functions. It interacts with the head domain of vinculin [119], reduces the affinity of β1-integrin for the extracellular matrix and increases the activity of RhoA [120]. These activities cause a rearrangement of the cellular actin, which results in the formation of membrane ruffles and an efficient uptake of the

bacterial cell. IpaB and IpaC form the translocator, a 25 Å pore in the host cell membrane, through which further effector proteins are transported [104], but they also exert additional functions. IpaB interacts with CD44 [121] and α5β1 integrin [122], which facilitates the bacterial uptake. It also activates caspase 1, leading to apoptosis in macrophages [123], and disrupts the vacuolar membrane, allowing *Shigella* cells to escape to the cytoplasm [124]. IpaC induces actin polymerization and formation of filopodia [125]. IpgB1 and IpgB2 mimic the effects of activated Ras-like GTPases. They modulate actin dynamics and lead to the formation of stress fibers and membrane ruffles [126]. IpgD dephosphorylates PIP2, resulting in the detachment of the cytoskeleton from the plasma membrane, thus facilitating the extension of membrane ruffles [127]. Finally VirA interacts with microtubules and destabilizes them, leading to the activation of Rac and membrane ruffling [128]. It has also been shown to facilitate the intracellular motility [129], although this might not be a direct effect of VirA [130]. Together, these effector proteins modulate the actin polymerization system, with the aim to remodel the actin cytoskeleton and the cell surface, to promote the uptake of a bacterial cell by phagocytosis into a normally non-phagocytic cell.

1.3.1 The pathogenic camouflage mechanism of *S. flexneri*

Intracellular pathogens are normally degraded by autophagy as described in chapter 1.1. However, *Shigella* developed a camouflage mechanism enabling it to avoid autophagic recognition and degradation. In 2005, the molecular basis of this camouflage mechanism has been discovered [46]. In a previous publication, it was shown that the effector protein IcsB is important for the pathogenicity of *Shigella* [131]. Upon infection of BHK or MDCK cells with *Shigella* IcsB deletion mutants, the bacteria initially proliferated normally. However, after four hours multiplication plateaued and the bacteria colocalized with GFP-LC3, an autophagosome specific marker, indicating that the cells were targeted for autophagic degradation. Additional experiments confirmed that the phenotype of the IcsB deletion mutant was associated with autophagy and that the cells were enclosed by autophagosomes. At the beginning of the infection no influence of the lack of IcsB on the ABM was observed. After two hours the bacteria became less motile, which may be caused by the trapping in autophagosomes. Complementation of IcsB expression restored the wild type phenotype [46].

In BHK cells, infected with *Shigella* IcsB deletion mutants, occasional localization of ATG5 and LC3 to one of the bacterial cell poles was observed, a distribution that is typical for the *Shigella* surface protein VirG. In a second experiment, the association of *Shigella* VirG deletion mutants with autophagy markers was barely detectable. In both cases the reconstitution of protein expression restored the wild type phenotype. Deletion mutants deficient in the expression of both IcsB and VirG showed the recovery of *Shigella* intracellular growth. Together, these results indicate that VirG is the target of autophagic recognition of *Shigella*. This hypothesis was tested by the infection of BHK cells with *E. coli* expressing VirG. The association of LC3 with these cells was observed, confirming the previous results [46].

In pulldown and immunoprecipitation experiments VirG was able to precipitate ATG5 from COS-7 and 293T cell lysates, but only if IcsB was not present. Increasing amounts of purified IcsB in the mixture resulted in decreased binding of ATG5 to VirG in a dose dependent manner, indicating a competitive binding of IcsB and ATG5 to VirG. The binding region of both proteins was located in between residues 320-433 of VirG [46].

Finally a model was presented, stating that the interaction of ATG5 and VirG targets the bacterial cell for autophagic degradation and might act as a seed for the isolation membrane, which engulfs the bacterium. The secretion of the effector protein IcsB constitutes the *Shigella* camouflage mechanism, as IcsB competitively binds to VirG with a higher affinity than ATG5, thus masking the bacterial cell and avoiding recognition by the host autophagy system (Figure 12).

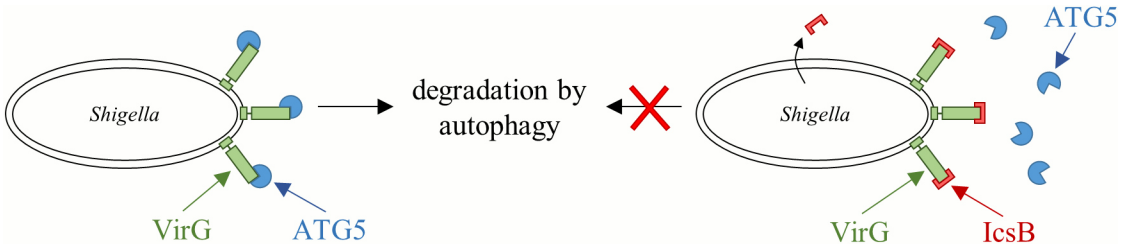


Figure 12: The *Shigella* outer membrane protein VirG (green) is bound by human ATG5 (blue), thus targeting the bacterial cell for autophagic degradation. As a countermeasure *Shigella* secretes the IcsB protein (red), which occupies the ATG5 binding site on VirG with a higher affinity than ATG5. The bacterial cell is not recognized by the innate immune response anymore and is free to move and proliferate inside the human cytoplasm.

In the next chapters the proteins VirG, IcsB and ATG5, which are involved in the autophagy evasion mechanism of *Shigella* are introduced in greater detail.

1.3.2 VirG

VirG is a *S. flexneri* outer membrane protein of 116 kDa and the target for autophagic recognition of *Shigella* cells by ATG5. It is composed of 1102 amino acids and contains three major domains. Residues 1-52 comprise an unusual signal peptide, responsible for the secretion to the periplasm by the sec-pathway. The passenger- or α-domain, comprising residues 53-758, is the soluble domain, which is presented on the cell surface. It has a MW of 78.0 kDa and a pI of 4.7. The β-core, comprising amino acids 759-1102, adopts a β-barrel fold and anchors VirG to the bacterial outer membrane [132]. An overview of the VirG domains is given in Figure 13 a. The *virG* gene is located on the *Shigella* virulence plasmid [100].

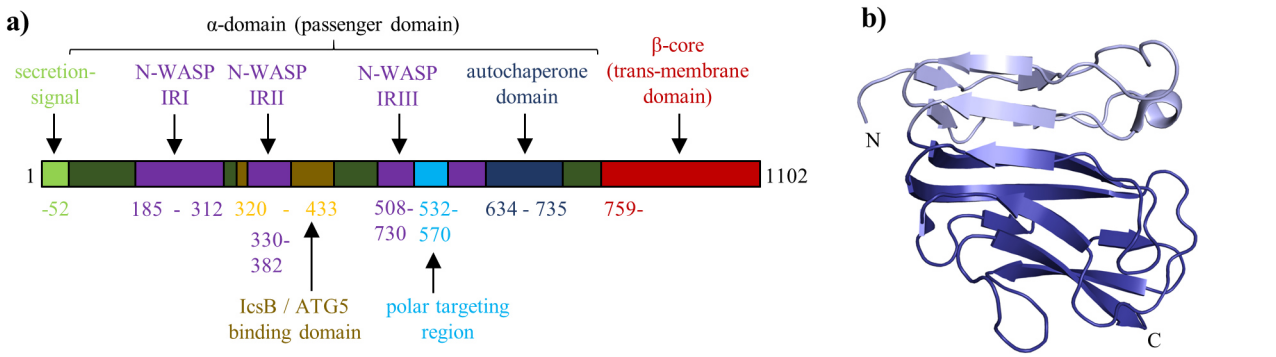


Figure 13: a) Annotated functional regions in VirG. b) Structure of the autochaperone domain from PDB-entry 3ML3. In the N-terminal part (light blue) it shows two turns of a right-handed parallel β-helix, while the C-terminal part (dark blue) consists of a β-sandwich [133].

VirG is a member of the type Va secretion system [134], also called an autotransporter protein or more precisely a self-associating autotransporter (SAAT), as it homo-oligomerizes [135], [136]. Already in the cytoplasm, it localizes to the old pole of the bacterial cell, where it is secreted by the sec-pathway to the periplasm [137], [138]. Once there, the signal peptide is removed by a special signal peptidase [139] and the β-core inserts into the outer membrane. Subsequently the passenger domain is passed through the β-core and is presented on the cellular surface [132]. The distribution of VirG observed in the cytoplasm is retained on the bacterial surface, where it forms a gradient with the highest concentration at the old cellular pole. The gradient formation is further enhanced by the activity of the specific protease IcsP, which cleaves VirG between positions 758 and 759, resulting in the release of the α-domain [140]. The presence of LPS is essential for the correct localization of VirG, as defects in the LPS result in abnormal distribution of VirG on the cellular surface [141], [142].

The α-domain, which is presented on the bacterial surface, contains a number of subdomains and stretches with annotated functions. Residues 634-735 form the autochaperone domain, which assists in the folding of the α-domain by forming a structured nucleus, once it is passed through the β-core [143]. A stable fragment of this domain has been identified by limited proteolysis and its structure was solved by X-ray crystallography [133] (Figure 13 b). A β-sandwich structure constitutes the C-terminal part of the autochaperone domain and the N-terminal part is formed by two turns of a right-handed β-helix. The stretch between residues 532-570 is called the polar targeting region. It is sufficient to localize VirG to the old pole of the bacterial cell, although the responsible molecular system and binding partners have not been identified yet [136]. A region comprising residues 320-433 binds both human ATG5 and *Shigella* IcsB and plays a role in the targeting of *Shigella* cells by autophagy and the bacterial camouflage mechanism as described in chapter 1.3.1.

The region comprising residues 117-332 is formed by a series of glycine-rich repeats (GRRs) [144]. Residues 103-433, which comprise the GRRs and additional N- and C-terminal stretches, are sufficient for the interaction with N-WASP *in vitro* [145]. For N-WASP interaction *in vivo* and actin tail formation residues 53-729, so essentially the whole α-domain, are necessary [146]. Recently the N-WASP binding region was narrowed down to three stretches: 1) N-WASP interaction region I (IRI) comprising residues 185-312, 2) IRII comprising residues 330-382 and 3) IRIII comprising residues 508-730. Amino acids V382, Y716 and D717 are of special importance for this interaction. It has also been shown that one molecule of N-WASP interacts with different molecules of VirG [135]. VirG binds the N-WASP-GBD in a manner similar to Cdc42 and thereby activates N-WASP [85], [89]. Activated N-WASP in turn activates the Arp2/3 complex that generates a propulsive force by forming the branched actin tail, typical for the intracellular motility of *Shigella* cells (chapter 1.2).

Residues 104-506 are essential for the interaction with vinculin, a protein that links actin filaments and focal adhesions [147]. Vinculin interacts with VASP and the presence of VASP could enhance actin polymerization [148]. It was also proposed that the actin filaments, which are bound to vinculin, facilitate the formation of new branches by the Arp2/3 complex [149]. However the interaction with vinculin is not essential for ABM [150]. Only recently, residues 507-758 were shown to interact with apyrase, a periplasmic *Shigella* enzyme with ATPase activity. Apyrase has been shown to be important for the unipolar localization of VirG and intercellular spread [151]. An overview of the different functional regions of the VirG α-domain is given in Figure 13 a.

Two different functions have been assigned to the VirG protein. The first is the function in intracellular ABM by the activation of N-WASP, already described in chapter 1.2. VirG is the only *Shigella* protein that is essential for ABM, as VirG expressing *E. coli* cells also show ABM *in vitro* [152]. Earlier in the infection cycle, when *Shigella* cells reside in the intestinal lumen, VirG acts as an adhesin binding to a still unknown receptor on the surface of human cells [108], which is its second function. VirG shows different proteolytic digestion patterns, depending on what function it exerts, indicating that the two states adopt different conformations [108]. Additionally VirG has an ATPase activity, whose biological role has yet to be identified [144].

1.3.3 IcsB and IpgA

The IcsB protein is the factor, which is responsible for the *Shigella* camouflage mechanism as described in chapter 1.3.1. IcsB is a 56.4 kDa protein that is coded upstream the *ipaBCDA*-operon, which is located in the pathogenicity island of the virulence plasmid. It comprises 494 amino acids with a calculated pI of 9.3. It is an effector protein, secreted *via* the *Shigella* TTSS, and is complexed by a specific chaperone as long as it resides inside the bacterial cytoplasm. In the case of IcsB, this chaperone is IpgA [131]. The structure of IcsB or homologues, like *Burkholderia spp.* BopA, is currently unknown.

Functional annotations exist for several regions of the IcsB protein. The first 15 amino acids represent a signal peptide, which is important for secretion by the TTSS. The chaperone binding domain (CBD), comprising residues 171-247, binds the specific chaperone IpgA [131]. Additionally a Rho-GTPase inactivation domain (RID) has been identified by a bioinformatics approach, which resides between residues 128-169 and 263-319. This domain is predicted to have a protease or acetyltransferase function and to inactivate small Rho-GTPases like Rho, Rac or Cdc42 [153]. Between residues 288-351, a cholesterol binding domain has been located [154]. In the UNIPROT database [155], a potential coiled coil domain, comprising residues 431-458, is annotated that has been identified by prediction software, e.g. COILS [156]. The predicted coiled coils might be involved in a so far unreported dimerization of the protein. An overview of the functional regions of IcsB is given in Figure 14.

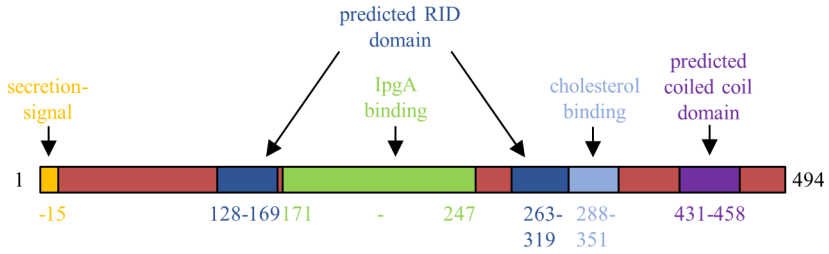


Figure 14: Annotated functional regions in IcsB.

IpgA is the specific chaperone of IcsB and a small protein of 15.1 kDa. It is composed of 129 amino acids and has a calculated pI of 4.8. It is coded directly downstream of *icsB* on the virulence plasmid [131]. Unlike typical chaperones that help in the folding of proteins, the function of IpgA is to keep the CBD of IcsB in an unfolded state. It also serves for the recognition of IcsB as a substrate for the TTSS. The inner diameter of the TTSS needle is about 20 Å [157]. Hence, secreted proteins have to pass the needle in an at least partially unfolded state. The already unfolded CBD inserts easily into the needle, while the tertiary structure of the remaining protein is removed in an ATP-dependent fashion, during the transport of the protein through the needle complex [158]. IpgA is removed in this process and stays in the bacterial cytoplasm, while IcsB refolds after its secretion into the host cytoplasm.

IcsB is required for *Shigella* pathogenicity [131]. It has a function in binding VirG, which comprises the *Shigella* camouflage mechanism, as described in chapter 1.3.1. The VirG binding regions in IcsB have not been identified yet. Secreted IcsB localizes to the bacterial surface [131] and, as recently described, helps in the recruitement of Toca-1 to the surface of *Shigella* cells [159]. Toca-1 is required for the activation of N-WASP, as described in chapter 1.2. Even though Toca-1 is essential for *Shigella* ABM [160], the localization to the bacterial surface is not, as deletion of IcsB does not lead to decreased intracellular mobility [131] and IcsB is not necessary for actin tail formation [159].

1.3.4 ATG5

ATG5 is the human protein that recognizes VirG and targets *Shigella* cells for autophagic degradation, as described in chapter 1.3.1. ATG5 is a 32.4 kDa protein, which is composed of 275 amino acids and has a calculated pI of 5.5. A second isoform has been reported that results from alternative splicing and begins with residue M80 [155]. ATG5 consists of two ubiquitin-like fold domains, UFD-1 and UFD-2, and an α-helical bundle region (HBR) [161]. An overview of the annotated domains of ATG5 is shown in Figure 15 a. At the beginning of this PhD work, structural information on ATG5 was limited to the X-ray structure of the *Saccharomyces cerevisiae* protein in complex with N-terminal fragments of Atg16 (PDB-entries 2DYM, 2DYO) (Figure 15 b). Residue K130 in the HBR is coupled to the C-terminal glycine of ATG12 *via* an isopeptide-bond in an ubiquitin-like conjugation reaction, involving both ATG7 and ATG10, as described in chapter 1.1. Residue T75 is a phosphorylation site of the MAP-Kinase p38. Phosphorylation in this location results in reduced autophagy [162]. Between residues T193

and T194, a calpain cleavage site has been located. Cleaved ATG5 is transported to mitochondria, which results in the release of cytochrome c and the activation of caspases, leading to apoptosis [163]. However, more recent studies couldn't confirm the previously reported calpain mediated cleavage of ATG5 [20], [164].

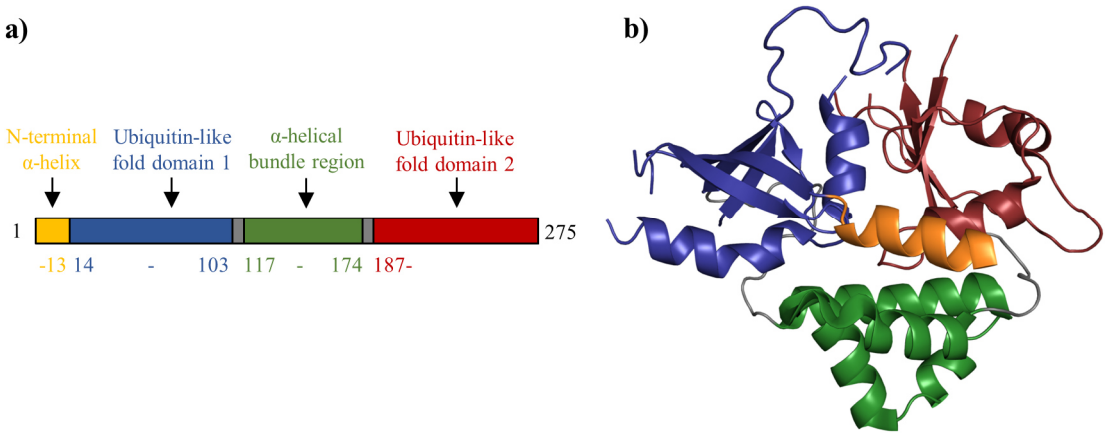


Figure 15: a) Annotated domains in ATG5. b) *S. cerevisiae* Atg5 structure from PDB-entry 2DYM. Atg5 consists of an N-terminal α-helix (orange), two ubiquitin-like fold domains (UFDs, blue and red), an α-helical bundle region (HBR, green) and two linker regions (grey).

ATG5 exerts a number of different functions in human cells. 1) ATG5 is a central regulator in autophagy. ATG16L1 and TECPR1 bind competitively to ATG5 by means of their ATG5 interaction motifs (AFIMs) [165]. Three stretches in ATG5, comprising residues D4-R15, P34-K51 and H241-T249, are implicated in these interactions (derived from PDB-entries 4GDK, 4TQ1). Binding with ATG16L1 is associated with autophagosomal elongation, as described in chapter 1.1. In contrast binding to TECPR1 is a prerequisite for the fusion of autophagosomes and lysosomes [166]. 2) ATG5 also regulates apoptotic cell death. The interaction of ATG5 with the death domain of FADD plays an important role in apoptosis, as depletion of ATG5 results in suppression of IFN-γ induced cell death, while overexpression of ATG5 results in the induction of cell death [36]. FADD is a protein that interacts with the death receptor Fas and transfers the death signal by binding and activation of procaspase 8 [167]. 3) ATG5 has been shown to interact with VirG, resulting in targeting of internalized bacterial cells by autophagy, as described in chapter 1.3.1.

2 Aims of this work

The objective of this work was a more detailed understanding of the molecular basis of the *Shigella* camouflage mechanism, using biochemical and structural biology methods. For this reason, the three involved proteins, VirG, IcsB and ATG5, first had to be produced as homogenous and soluble protein in high purity and quantity, to be suitable for crystallization experiments. Then the structures of the three individual proteins, as well as the complex structures of VirG/IcsB and VirG/ATG5, should be solved by X-ray crystallography to characterize the interaction surfaces. Using these results, a reporter system should be designed, to use a high throughput screening method for the search of a low MW compound that acts as an inhibitor of the camouflage mechanism, by disrupting the VirG/IcsB interaction, while leaving the VirG/ATG5 interaction undisturbed. This compound would have reestablished the host cell's pathogen defense systems and induce autophagy and degradation of *Shigella*, hence it could be used as a lead structure for the development of a new therapeutic treatment, specific against *Shigella* infections. Biophysical and biochemical methods should be used for the further characterization of the proteins, complexes and interactions involved in the camouflage mechanism, e.g. SAXS for the characterization of the proteins in solution and different interaction methods for the determination of dissociation and inhibitory constants.

As a side project, the molecular mechanism for the inhibitory effect of the phosphorylation of ATG5 residue T75 on autophagy, should be investigated. To this end, the structures of a phosphomimetic ATG5 mutant T75E and a control mutant T75A, should be elucidated and compared to the structure of the wild type protein. These results should be complemented with a biophysical characterization of the ATG5 variants and interaction studies with ATG16L1.

3 Materials and Methods

3.1 Materials

3.1.1 Buffers, solutions and media

Table 1: Buffers used in this work.

Name	Formulation
PBS buffer (10x)	1.37 M NaCl 27 mM KCl 14.4 g/l Na_2HPO_4 2.4 g/l KH_2PO_4 (≙ 120 mM Phosphate pH 7.4) adjust pH to 7.4, using 10 M NaOH sterile-filter, using 0.2 µm filter, store at room temp.
PBS-T buffer	0.1% (v/v) Tween20 in 1x PBS store at room temperature
EDTA	0.5 M EDTA adjust pH to 8.0 to dissolve sterile-filter, using 0.2 µm filter, store at room temp.
DNA loading buffer (10x)	20% (w/v) Ficoll 400 0.25% (w/v) Bromphenol blue 0.25% (w/v) Xylene cyanol in 0.5 M EDTA pH 8.0 store at -20°C
TAE buffer (50x)	2 M TRIS base 1 M Acidic acid 50 mM EDTA autoclave, store at room temperature
SDS loading buffer (2x)	100 mM TRIS pH 6.8 4% (w/v) SDS 20% (v/v) Glycerol 200 mM DTT 0.2% (w/v) Bromphenol blue store at -20°C
Laemmli buffer (10x)	0.25 M TRIS pH 8.5 1.92 M Glycine 1% (w/v) SDS store at room temperature
Coomassie Brilliant Blue staining solution	0.1% (w/v) Coomassie brilliant blue R250 40% (v/v) Ethanol 10% (v/v) Acidic acid store at room temperature
Destaining solution	40% (v/v) Ethanol 10% (v/v) Acidic acid store at room temperature
CD buffer	1.378 g/l $Na_2HPO_4 \cdot 2H_2O$ 0.352 g/l $NaH_2PO_4 \cdot 2H_2O$ (≙ 10 mM Phosphate pH 7.4), prepare freshly

Lysine methylation buffer	50 mM HEPES pH 7.5 250 mM NaCl prepare freshly
Lysine methylation SEC buffer	20 mM TRIS pH 7.5 200 mM NaCl prepare freshly
BLI regeneration buffer	25 mM Glycine adjust pH to 2.5
Transfer buffer	25 mM TRIS pH 8.0 192 mM Glycine 10% (v/v) Methanol store at room temperature
TBS buffer (10x)	100 mM TRIS pH 7.5 1.5 M NaCl sterile-filter, using 0.2 µm filter, store at room temp.
TBS-T buffer	0.1% (v/v) Tween20 in 1x TBS store at room temperature
Blocking solution	5% (w/v) skim milk powder in TBS-T store in fridge
Staining solution	10 ml alkaline phosphatase buffer 66 µl NBT 33 µl BCIP
Test-expression lysis buffer	50 mM TRIS pH 8.0 200 mM NaCl 5 mM BME 1 mM PMSF prepare freshly
Test-expression washing buffer	50 mM TRIS pH 8.0 200 mM NaCl 5 mM BME 10 mM Imidazole prepare freshly
VirG lysis buffer	50 mM HEPES pH 7.0 300 mM NaCl 5 mM BME prepare freshly
IcsB lysis buffer	50 mM HEPES pH 7.4 200 mM NaCl 5 mM BME prepare freshly
ATG5 lysis buffer	50 mM HEPES pH 7.0 500 mM NaCl 5% Glycerol 5 mM BME prepare freshly
washing buffer	lysis buffer with 30 mM Imidazole, prepare freshly

elution buffer	lysis buffer with 250 mM Imidazole, prepare freshly
regeneration buffer	lysis buffer with 500 mM Imidazole, prepare freshly
SourceQ low salt buffer	50 mM TRIS pH 8.0, 5 mM DTT sterile-filter, using 0.2 µm filter, degas, prepare freshly
SourceQ high salt buffer	50 mM TRIS pH 8.0, 1 M NaCl, 5 mM DTT sterile-filter, using 0.2 µm filter, degas, prepare freshly
SourceS low salt buffer	50 mM MES pH 6.0, 5 mM DTT sterile-filter, using 0.2 µm filter, degas, prepare freshly
SourceS high salt buffer	50 mM MES pH 6.0, 1 M NaCl, 5 mM DTT sterile-filter, using 0.2 µm filter, degas, prepare freshly
SEC buffer	lysis buffer with 5 mM DTT instead of BME sterile-filter, using 0.2 µm filter, degas, prepare freshly

Table 2: Media used in this work.

Name	Formulation
LB medium	10 g/l Tryptone 5 g/l Yeast extract 7 g/l NaCl for solid medium add 16 g/l agar autoclave, store at room temperature
TB salts (10x)	0.17 M KH_2PO_4 0.72 M K_2HPO_4 autoclave, store at room temperature
TB medium	12 g/l Tryptone 24 g/l Yeast extract 0.4% (v/v) Glycerol fill up to 900 ml autoclave, store at room temperature add 100 ml TB salts before use
Mg solution (100x)	1 M $MgCl_2$ 1 M $MgSO_4$ autoclave, store at room temperature
Glucose solution (100x)	2 M Glucose sterile-filter, using 0.2 µm filter, store at room temperature
SOC medium	20 g/l Tryptone 5 g/l Yeast extract 10 mM NaCl 2.5 mM KCl autoclave before use add 10 ml/l Mg solution, 10 ml/l glucose solution, adjust pH to 7.0 using 10 M NaOH sterile-filter, using 0.2 µm filter, store at -20°C

Table 3: Antibiotics used in this work.

Name	Formulation
Ampicillin (1000x)	100 g/l Ampicillin sterile-filter, using 0.2 µm filter, store at -20°C
Kanamycin (1000x)	30 g/l Kanamycin sterile-filter, using 0.2 µm filter, store at -20°C
Chloramphenicol (1000x)	34 g/l Chloramphenicol in 70% (v/v) ethanol store at -20°C

Table 4: Buffers and media used for the ESPRIT system in the laboratories of Darren Hart at the EMBL in Grenoble.

Name	Formulation
$MgSO_4$ solution	9 g/l $MgSO_4 \cdot 7H_2O$ autoclave, store at room temperature
Hogness modified freezing medium (HMFM, 10x)	63 g/l K_2HPO_4 18 g/l KH_2PO_4 9 g/l $Na_3Citrate$ 44% (v/v) Glycerol autoclave add 50 ml $MgSO_4$ solution sterile-filter, using 0.2 µm filter, store at room temp.
Biotin solution	20 mM Biotin dissolve in half of the final volume add 1 M NaOH until the solution clears up fill up to final volume sterile filter, using 0.2 µm filter, store in fridge
Denaturation solution	0.5 M NaOH 1.5 M NaCl store at room temperature
Neutralization solution	1 M TRIS 1.5 M NaCl adjust pH to 7.5 using HCl store at room temperature
Saline sodium citrate (SSC) buffer (20x)	0.3 M $Na_3Citrate$ 3 M NaCl adjust pH to 7.0 using HCl store at room temperature
Gomori buffer (10x)	40.96 g/l Na_2HPO_4 25.37 g/l NaH_2PO_4 (≙ 0.5 M Phosphate pH 7.0) sterile-filter, using 0.2 µm filter, store at room temp.
Spheroplast formation buffer	20 mM TRIS pH 8.0 300 mM NaCl 20% (w/v) Sucrose 0.1% (w/v) Lysozyme store in fridge

Lysis buffer	10 mM TRIS pH 7.0 0.5% (v/v) Brij 0.005% (v/v) Benzonase store in fridge
Washing buffer	300 mM NaCl 5 mM Imidazole in 1x Gomori buffer store in fridge
Elution buffer	300 mM NaCl 300 mM Imidazole in 1x Gomori buffer store in fridge
Blocking solution	5% (w/v) whole milk powder in PBS-T store in fridge

3.1.2 Bacterial strains

Table 5: Bacterial strains used in this work.

Name	Description	Supplier
DH5α	F^- λ^- *end*A1 *gln*V44 *thi*-1 *rec*A1 *rel*A1 *gyr*A96 Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *hsd*R17 (r_K^- m_K^+) *pho*A *sup*E44	Thermo Fisher Scientific Inc.
Rosetta 2 (DE3)	F^- *omp*T *hsd*S_B(r_B^- m_B^-) *gal dcm* (DE3) pRARE2 (Cam^R)	Merck KGaA

3.1.3 Plasmids

Table 6: Plasmids used in this work.

Identifier	Description	Source
pSA010	pGEX_GST_Thr_TEV_mcs	own lab
pSA035	pET15b_His_TEV_mcs	own lab
pSA052	pCOLA_Duet_His_MBP_TEV_mcs1_mcs2_S	own lab
pVP006	pCOLA_Duet_Strep_TEV_mcs1_mcs2_S	own lab
pVP008	pCOLA_Duet_Strep_TEV_mcs1_mcs2_S	own lab
pLS063	pESPRIT002_His_TEV_mcs_BAP	received from Darren Hart, EMBL
pLS064	pACYC_Flag_TEV_mcs	received from Darren Hart, EMBL
pAA054	pCOLA_Duet_ZZ_His_TEV_mcs1_mcs2_S	own lab

3.1.4 Oligonucleotides

The oligonucleotides used in this work are listed in the appendix chapter A1.

3.1.5 Ordered Genes and peptides

A synthetic human Toca-1 gene, which was codon optimized for *E. coli* expression, was ordered from Eurofins Genomics GmbH. An ATG16L1 peptide, comprising residues 11-43, was ordered from an in house peptide synthesis service.

3.1.6 Kits

Table 7: Kits used in this work.

Name	Supplier
QIAquick PCR Purification Kit	Qiagen N.V.
QIAquick Gel Extraction Kit	Qiagen N.V.
QIAprep Spin Miniprep Kit	Qiagen N.V.

3.1.7 Chemicals and Biochemical

Chemicals were bought in quality p.A. or similar from the following companies: Carl Roth, Honeywell Riedel-de Haën, J.T. Baker, Merck KGaA, SERVA, Sigma-Aldrich and VWR International.

Table 8: Enzymes used in this work.

Name	Supplier
2x Red PCR Master Mix	P.J.K. GmbH
KOD Hot Start DNA Polymerase	Merck KGaA
T4 DNA Ligase	Thermo Fisher Scientific Inc.
T4 Polynucleotide Kinase	Thermo Fisher Scientific Inc.
Antarctic phosphatase	New England Biolabs
DNase I	F. Hoffmann-La Roche AG
Restriction Endonucleases: *Not*I-HF, *Kpn*I-HF	New England Biolabs
Lysozyme	Sigma-Aldrich
TEV protease	own lab

Table 9: Standards used in this work.

Name	Supplier
GeneRuler DNA Ladder Mix	Thermo Fisher Scientific Inc.
Unstained Protein Molecular Weight Marker	Thermo Fisher Scientific Inc.
ColorPlus Prestained Protein Marker	New England Biolabs

Table 10: Column materials used in this work.

Name	Supplier
Superdex 200 pg	GE Healthcare
Superdex 200 GL	GE Healthcare
Superdex 75 GL	GE Healthcare
SOURCE 15Q	GE Healthcare
SOURCE 15S	GE Healthcare
Ni Sepharose 6 Fast Flow	GE Healthcare
Strep-Tactin Superflow high capacity	IBA GmbH
Strep-Tactin Macroprep	IBA GmbH
Glutathion Sepharose 4 Fast Flow	GE Healthcare

3.1.8 Crystallization screens

Table 11: Crystallization screens used in this work.

Name	Supplier
Additive Screen	Hampton Research
$AmSO_4$	Qiagen N.V.
Anions	Qiagen N.V.
Cations	Qiagen N.V.
Index	Hampton Research
JB Classic HTS II	Jena Bioscience GmbH
JCSG+	Qiagen N.V.
JCSG Core I	Qiagen N.V.
JCSG Core II	Qiagen N.V.
JCSG Core III	Qiagen N.V.
JCSG Core IV	Qiagen N.V.
MBClass	Qiagen N.V.
MIDAS	Molecular Dimensions Ltd.
Morpheus	Molecular Dimensions Ltd.
PEGs	Qiagen N.V.
PEGs II	Qiagen N.V.
PGA-LM	Molecular Dimensions Ltd.
pHClear	Qiagen N.V.
Protein Complex	Qiagen N.V.
Silver Bullets	Hampton Research

3.1.9 Laboratory equipment

Table 12: Laboratory equipment used in this work.

Name	Function	Supplier
BP2100S, I2000D	Balance	Sartorius AG
XB6200D	Balance	Precisa Gravimetrics AG
XS204	Balance	Mettler-Toledo International Inc.
Variomag Mono Direct	Magnetic stirrer	H+P Labortechnik GmbH
MR 2000/2002/3001/ Hei-Standard	Magnetic stirrer	Heidolph Instruments
Thermomixer comfort/5436/5437	Thermo mixer	Eppendorf AG
Nutating Mixer	Nutating mixer	VWR International
Vortex-Genie 2	Vortex mixer	Scientific Industries Inc.
Sorvall RC 6 (Plus) Heraeus Multifuge X3R Heraeus Fresco 21	Centrifuge	Thermo Fisher Scientific Inc.
5417R/5810R	Centrifuge	Eppendorf AG
Micro Star 17(R)	Centrifuge	VWR International
R-963	Microwave	Sharp K.K.
Milli-Q Advantage A10	Water purification system	Merck KGaA
Herasafe	Cleanbench	Thermo Fisher Scientific Inc.

T-Personal, T-Professional, T3000	Thermocycler	Biometra GmbH
CFX96 Real-Time PCR Detection System	Realtime PCR	Bio-Rad Laboratories
PowerPac 300/Basic	Power Supply	Bio-Rad Laboratories
Mini-PROTEAN 3 Cell/Tetra Cell	SDS-gel running system	Bio-Rad Laboratories
Trans-Blot Turbo	Western blot transfer system	Bio-Rad Laboratories
(Wide) Mini-Sub Cell GT Cell	Agarose gel running system	Bio-Rad Laboratories
Gel Logic 212 Pro	Gel documentation system	Carestream Health Inc.
UV Transilluminator	UV transilluminator	Intas Science Imaging Instruments GmbH
Heraeus microbiological incubator	Incubator	Thermo Fisher Scientific Inc.
Multitron Standard	Incubation shaker	Infors AG
Bandelin Sonopuls HD 200	Ultrasonic homogenizer	BANDELIN electronic GmbH & Co. KG
TS Series Benchtop	Homogenizer	Constant Systems Ltd.
EmulsiFlex-C3	Homogenizer	Avestin Inc.
ÄKTApurifier	Liquid chromatography system	GE Healthcare
CO8000 Cell Density Meter	Cell density meter	Biochrom Ltd.
ND-1000	UV-Vis spectrophotometer	Thermo Fisher Scientific Inc.
Ultrospec 3000	UV-Vis spectrophotometer	Pharmacia
J-815 Circular Dichroism Spectrometer	CD-spectrometer	JASCO Inc.
MicroCal VP-ITC	ITC	GE Healthcare
Octet RED96	BLI	Pall Corp.
Monolith NT.115	MST	Nanotemper Technologies GmbH
Formulator	Liquid handling system	Formulatrix Inc.
Rock Imager 182/1000	Imaging system	Formulatrix Inc.
Honeybee 961	Protein crystallization robot	Digilab Inc.
OryxNano	Protein crystallization robot	Douglas Instruments Ltd.
SteREO Discovery.V8 Stemi SV 8	Stereomicroscope	Carl Zeiss AG
SZ40	Stereomicroscope	Olympus
Micromax-007 HF	X-ray generator	Rigaku Corp.
Saturn 944+ R-AXIS IV++	X-ray detector	Rigaku Corp.

Table 13: Laboratory equipment used for the ESPRIT system in the laboratories of Darren Hart at the EMBL in Grenoble.

Name	Function	Supplier
QFill2	Microplate dispenser	Genetix
HiGro	Plate incubation shaker	GeneMachines
	Picking and gridding robot	Kbiosystems robotic solutions
Genesis Workstation 200	Liquid handling system	Tecan Group AG
Typhoon Trio Variable Mode Imager	Fluorescence scanner	GE Healthcare

3.2 Methods

3.2.1 Cloning

3.2.1.1 Polymerase chain reaction (PCR), 2 step PCR, Colony PCR

For the amplification of a DNA fragment for cloning, 0.5 µl template DNA, 5 µl *KOD* buffer (10x), 5 µl dNTPs (2 mM each), 3 µl $MgSO_4$ (25 mM), 2 µl forward primer (10 pmol/µl), 2 µl reverse primer (10 pmol/µl), 1 µl *KOD* DNA polymerase and 31.5 µl H_2O were mixed and the following PCR program was run:

1. 95°C, 2 min
2. 95°C, 30 sec
3. 55°C, 15 sec
4. 70°C, 20 sec/1000 bp → 2. (40x)
5. 70°C, 10 min
6. 10°C, hold

DNA loading buffer (10x) was added and the reaction mixture was loaded on a 0.8% agarose gel. The correct band was excised and the DNA was purified, using the QIAquick Gel Extraction Kit.

To perform a 2-step PCR, the two PCR reactions for the first step were run according to the protocol above. 0.5 µl of the two reactions were used as template to set up a new PCR reaction for the second step, which was again run according to the protocol above. The DNA was loaded on an agarose gel and purified, as described above.

For a colony PCR, 10 µl 2x Red PCR Master Mix, 2 µl forward primer, 2 µl reverse primer and 6 µl H_2O were mixed. An Eppendorf tube with 0.5 ml of LB-medium was prepared. A colony from a LB-agar plate was picked and transferred to both the PCR tube and the Eppendorf tube. The Eppendorf tube was incubated at 37°C. The following PCR program was run on the PCR tube:

1. 95°C, 5 min
2. 95°C, 30 sec
3. 55°C, 30 sec
4. 70°C, 1 min/1000 bp → 2. (26x)
5. 70°C, 10 min
6. 10°C, hold

The PCR reaction mixture was loaded directly on a 0.8% agarose gel and checked for a PCR product of the correct size. If the clone was positive, 5 ml of LB in a test-tube were inoculated with 200 µl of culture from the Eppendorf tube and incubated at 37°C, 130 rpm overnight.

3.2.1.2 Restriction enzyme digestion

≤ 10 µg plasmid or PCR fragment, 5 µl NEBuffer 4 (10x), 1 µl *Not*I-HF and 1 µl *Kpn*I-HF were mixed, filled up to 50 µl with H_2O and incubated at 37°C for 1 h. To dephosphorylate a plasmid, 5 µl antarctic phosphatase buffer and 2 µl antarctic phosphatase were added and incubated at 37°C for 30 min, then at 65°C for 5 min. DNA loading buffer (10x) was added and the reaction mixture was loaded on a 0.8% agarose gel. The correct band was excised and the DNA was purified, using the QIAquick Gel Extraction Kit.

3.2.1.3 Ligation

0.5 µl plasmid (digested, dephosphorylated), 4 µl insert (digested PCR fragment), 2 µl T4 DNA ligase buffer, 1 µl T4 DNA ligase and 12.5 µl H_2O were mixed and incubated at room temperature for 30 min.

3.2.1.4 Transformation

10 µl of ligation reaction mixture or 1 µl of purified plasmid was added to one aliquot of competent cells and incubated on ice for 30 min. The cells were heat shocked at 42°C for 90 sec, incubated on ice for 2 min and 1 ml of SOC medium was added. The cells were incubated at 37°C, 700 rpm for 1 h, then a part or all of them were plated on LB-agar and incubated at 37°C overnight. For cotransformations 2 µl of each purified plasmid were added and all of the cells were plated.

3.2.1.5 Plasmid sequencing

500 ng plasmid (concentration ≤ 500 ng/µl, volume ≤ 5 µl) and 1 µl primer (10 pmol/µl) per sequencing reaction were sent to an in house sequencing service to obtain the sequence of a plasmid.

3.2.1.6 Storage of bacterial strains

0.5 ml of overnight culture grown to stationary phase and 0.5 ml 86% glycerol were mixed briefly, using a vortex mixer and stored at -80°C (also possible at -20°C for short periods of time).

3.2.2 Protein production and purification

3.2.2.1 Small scale test-expressions

Small scale test-expressions were performed to determine the best expression conditions for the production of a certain protein of interest. Usually different temperatures, IPTG concentrations and expression times were tested, but also other conditions like the type of culture medium and the expression strain could have been varied. A large culture of an *E. coli* strain, expressing the protein of interest, in TB was inoculated from an overnight preculture to OD 0.03 and grown at 37°C, 130 rpm to OD 0.6-0.8. At this point the culture was split into a number of smaller cultures, each containing 30 ml of volume, and induced, using two different IPTG concentrations (50 and 200 µM). The cultures were further incubated at two different temperatures (20°C and 37°C), 130 rpm overnight. A first 10 ml sample was taken 3 h after induction, a second 10 ml sample after overnight expression.

The samples were centrifuged at 4°C, 5000 rpm for 5 min and the cell pellets were resuspended in 1 ml of lysis buffer each. The cells were lysed by sonication at 4°C, three times 10 seconds at 50% amplitude. 20 µl samples were taken and centrifuged at 4°C, 13000 rpm for 10 min. The pellet was boiled in 20 µl SDS loading buffer as a sample of the insoluble proteins. The rest of the lysed cells was centrifuged at 4°C, 13000 rpm for 30 min and the supernatant was given to 20 µl of the affinity beads, matching the affinity tag of the protein of interest. They were incubated at 4°C for 30 min for protein binding. The beads were spun down at 4°C, 2000 rpm for 1 min and washed twice in 1 ml of washing buffer. Finally the beads were boiled in 20 µl SDS loading buffer as a sample of the purified soluble protein.

All protein samples were analyzed on an SDS-gel. The expression condition, yielding most purified soluble protein, was identified and used in later large scale expressions.

3.2.2.2 Large scale protein expression and purification

Large scale expressions and purifications were performed to produce pure protein for crystallization setups and biochemical studies. Depending on the expected protein quantity, 0.5 l up to 16 l of TB were inoculated from an overnight preculture to OD 0.03 and grown at 37°C, 130 rpm to OD 0.6-0.8. If the best expression condition was at 20°C, the temperature was switched at OD 0.4. The cultures were induced, using the IPTG concentration determined in test-expressions (usually 50 µM) and grown for the determined period of time (usually overnight). The cultures were spun down at 4°C, 5000 rpm for 10 min, the pellets were weighed and resuspended 5·x ml of lysis-buffer per x g of pellet. 1 mM PMSF, 5 mM $MgCl_2$, 50 µg per liter of culture DNase and 5 mg per liter of culture lysozyme were added and the cells were lysed twice, using a homogenizer. The lysate was centrifuged at 4°C, 16000 rpm for 45 min.

For a His-tagged protein purification 20 mM imidazole were added to the cleared lysate. His-affinity beads were prepared by washing with 20 column volumes (CV) H_2O and 20 CV lysis buffer and added to the cleared lysate. The amount of beads was calculated, so that they were slightly overloaded, according to the expected protein quantity. The lysate was stirred at 4°C for 1 h for protein binding. The beads were loaded into a column and washed, using 50 CV of washing buffer. The protein was eluted from the beads with 10 CV of elution buffer before the beads were regenerated by washing with 10 CV regeneration buffer, 20 CV H_2O, 10 CV 5 M guanidinium chloride, 20 CV H_2O and 10 CV 20% EtOH (for storage).

For a Strep-tagged protein purification the Strep-affinity beads were prepared by washing with 20 CV of lysis buffer and added to the cleared lysate. The lysate was stirred at 4°C for 1 h for protein binding. The beads were loaded into a column and washed, using 50 CV of lysis buffer. The protein was eluted with four times 1 CV of lysis buffer, supplemented with 2.5 mM D-desthiobiotin. The beads were regenerated by washing with 10 CV of lysis buffer, supplemented with 1 mM HABA, and 20 CV of lysis buffer (also for storage). If the HABA couldn't be completely removed, a high pH washing step was added.

For a GST-tagged protein purification the lysis buffer was replaced by PBS buffer, as with PBS usually higher protein yields were obtained. The GST-affinity beads were prepared similar to His-affinity beads and added to the cleared lysate. The lysate was stirred at 4°C for 1 h for protein binding. The beads were loaded into a column and washed, using 50 CV PBS supplemented with 5 mM BME. The protein was eluted with four times 1 CV of PBS supplemented with 5 mM BME and 20 mM GSH (pH adjusted). Then the beads were regenerated similar to His-affinity beads.

If the affinity tag was to be cleaved off, 1-2% (w/w) TEV protease were added to the elution fraction and incubated at 4°C overnight. Samples of all affinity purification steps were analyzed on an SDS-gel and the protein concentrations in the elution fractions were determined spectrophotometrically.

The buffers for all following chromatography steps were 0.2 µm filtered and degassed. For ion exchange chromatography (IEC), the column was prepared by washing with 2 CV of H_2O and repeated low salt, high salt washing steps of 0.5 CV. Finally the column was equilibrated in low salt buffer. The protein was diluted to about 50 mM NaCl with low salt buffer, 0.2 µm filtered and loaded onto the column. The column was washed with 2 CV of buffer at 50 mM NaCl before a gradient to 750 mM NaCl over 20 CV was applied. The elution from the gradient was fractionated and analyzed on an SDS-gel to identify the protein containing fractions, which were pooled. The column was cleaned by washing with 5 CV high salt and 5 CV low salt buffer, before the washing program from the beginning was repeated, and it was stored in 20% EtOH.

For size exclusion chromatography (SEC) the column was prepared by washing with 1 CV of H_2O and equilibrated with 1 CV of SEC buffer. The protein was concentrated to 1-2% of the CV, using an appropriate Vivaspin concentrator. It was centrifuged at 4°C, 13000 rpm for 10 min to remove particles and loaded onto the column. 1 CV of SEC buffer was run over the column and the elution was fractionated and analyzed on an SDS-gel to identify the protein containing fractions, which were pooled. The column was washed with 1 CV H_2O and 1 CV 20% EtOH for storage. The protein was concentrated to 10-20 mg/ml before it was aliquoted, frozen in liquid nitrogen and stored at -80°C for further use.

3.2.3 Lysine methylation

Lysine methylation is a fast and easy method to modify the surface of a protein if it is not possible to identify initial crystallization conditions. The protein sample was dialyzed against 4 l lysine methylation buffer at 4°C overnight. It was diluted to a concentration of 1 mg/ml, 20 µl 1 M dimethylaminoborane complex (ABC) and 40 µl 1 M formaldehyde per ml of protein solution were added and it was incubated at 4°C for 2 h. Another 20 µl 1 M ABC and 40 µl 1 M formaldehyde per ml of protein solution were added and it was incubated at 4°C for another 2 h. Finally 10 µl 1 M ABC per ml of protein solution was added and the mixture was incubated at 4°C overnight. The protein was centrifuged at 4°C, 13000 rpm for 10 min to remove precipitate and separated on a SEC column in lysine methylation SEC buffer. This protocol was adapted from [168].

3.2.4 Protein characterization

3.2.4.1 Western blot

Western blotting was used to detect a certain protein or affinity tag in a complex mixture with high sensitivity. An SDS-gel was equilibrated in transfer buffer for 10 min and a PVDF-membrane was first activated in 100% MeOH for 5 min, then also soaked in transfer buffer for 10 min. The blotting device was assembled, first on the anode a Whatman paper soaked in transfer buffer, then the membrane, the gel and on top, in contact with the cathode, another transfer buffer soaked Whatman paper. The gel was blotted onto the membrane, using 25 V, 1 A for 30 min. The membrane was incubated in blocking solution at room temperature for 1 h. It was washed in TBS-T three times for 5 min, before it was incubated with the primary antibody (mouse anti-GST, 1:1000) in TBS-T at room temperature for 1 h. The membrane was washed again in TBS-T three times for 5 min and in TBS for 15 min before it was incubated with the secondary antibody (goat anti-mouse IgG alkaline phosphatase conjugate, 1:5000) at room temperature for 1 h. It was washed again in TBS-T three times for 5 min and in TBS for 15 min. Finally the blot was stained by adding staining solution. The staining was stopped by adding H_2O, then the membrane was washed in H_2O and dried.

3.2.4.2 Thermofluor-assay

The thermofluor-assay is a method to determine the melting temperature of a protein. It uses the SYPRO Orange dye, which binds to hydrophobic regions of proteins. In solution the fluorescence of the dye is quenched, but upon binding its fluorescence signal can be detected at 570 nm. The protein and the dye are mixed in a certain buffer condition and heated up slowly, starting at 10°C and finishing at 95°C, while the fluorescence is monitored by a real time PCR machine. Usually at low temperatures the fluorescence is low, as the protein is folded and mostly hydrophilic amino acids cover the surface. The protein unfolds at the melting temperature and the fluorescence signal increases, because the hydrophobic core of the protein is exposed and the dye binds to it. Upon further heating the fluorescence signal decreases again, as the protein aggregates and the hydrophobic residues are not accessible anymore. The melting point of the protein is defined as the inflection point of the fluorescence curve (Figure 16). By screening different buffer conditions, the optimal buffer for a certain protein can be identified. This buffer stabilizes the protein the most, resulting in the highest melting temperature.

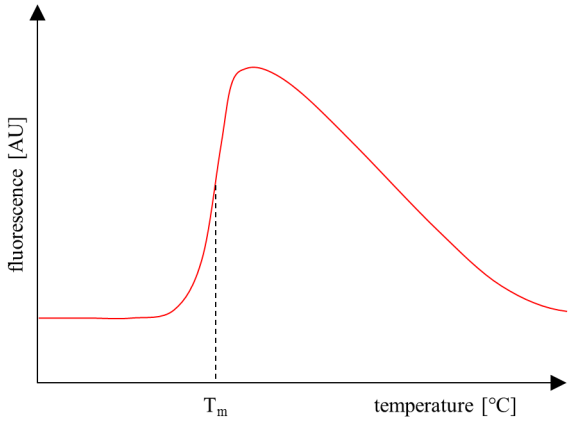


Figure 16: A typical melting curve of a protein in a thermofluor-assay. The SYPRO Orange dye is used that binds to hydrophobic residues. In solution its fluorescence is quenched, but upon binding its fluorescence signal can be detected at 570 nm. At low temperatures the dye does not bind to the protein and the fluorescence is low. At the melting temperature T_m the protein unfolds, the dye binds to the exposed residues from the hydrophobic core and the fluorescence signal rises. At even higher temperatures the protein aggregates, the hydrophobic residues are not accessible anymore and the fluorescence signal slowly decreases. T_m is defined as the first inflection point of the fluorescence curve.

First a test was run to determine the optimal concentrations of the protein and the dye. Dilutions of the protein to 0.5, 1 and 2 mg/ml, and of the dye to 100x, 50x and 10x from a 5000x stock (from supplier, unknown absolute concentration) were prepared. Every combination of protein- to dye-concentration was tested. 40 µl protein buffer, 5 µl protein and 5 µl SYPRO Orange were mixed for all the conditions and the experiment was run on a realtime PCR machine, as described above. From the appearance of the fluorescence curves, the best condition was selected. The curve should have had a shape as described above and a sufficient amplitude. Usually 2 mg/ml of protein and 50x or 100x SYPRO Orange were identified to yield the best results.

Table 14: Thermofluor-assay buffer and additive screen used in this work.

	1	2	3	4	5	6	7	8	9	10	11	12
A	50mM Acetate pH 4.0	50mM Acetate pH 4.4	50mM Citrate pH 5.0	50mM Citrate pH 5.4	50mM MES pH 6.0	50mM MES pH 6.4	50mM HEPES pH 7.0	50mM HEPES pH 7.4	50mM TRIS pH 8.0	50mM TRIS pH 8.4	50mM CHES pH 9.0	50mM CHES pH 9.4
B	50mM Acetate pH 4.0 100mM NaCl	50mM Acetate pH 4.4 100mM NaCl	50mM Citrate pH 5.0 100mM NaCl	50mM Citrate pH 5.4 100mM NaCl	50mM MES pH 6.0 100mM NaCl	50mM MES pH 6.4 100mM NaCl	50mM HEPES pH 7.0 100mM NaCl	50mM HEPES pH 7.4 100mM NaCl	50mM TRIS pH 8.0 100mM NaCl	50mM TRIS pH 8.4 100mM NaCl	50mM CHES pH 9.0 100mM NaCl	50mM CHES pH 9.4 100mM NaCl
C	50mM Acetate pH 4.0 200mM NaCl	50mM Acetate pH 4.4 200mM NaCl	50mM Citrate pH 5.0 200mM NaCl	50mM Citrate pH 5.4 200mM NaCl	50mM MES pH 6.0 200mM NaCl	50mM MES pH 6.4 200mM NaCl	50mM HEPES pH 7.0 200mM NaCl	50mM HEPES pH 7.4 200mM NaCl	50mM TRIS pH 8.0 200mM NaCl	50mM TRIS pH 8.4 200mM NaCl	50mM CHES pH 9.0 200mM NaCl	50mM CHES pH 9.4 200mM NaCl
D	50mM Acetate pH 4.0 500mM NaCl	50mM Acetate pH 4.4 500mM NaCl	50mM Citrate pH 5.0 500mM NaCl	50mM Citrate pH 5.4 500mM NaCl	50mM MES pH 6.0 500mM NaCl	50mM MES pH 6.4 500mM NaCl	50mM HEPES pH 7.0 500mM NaCl	50mM HEPES pH 7.4 500mM NaCl	50mM TRIS pH 8.0 500mM NaCl	50mM TRIS pH 8.4 500mM NaCl	50mM CHES pH 9.0 500mM NaCl	50mM CHES pH 9.4 500mM NaCl
E	empty											
F	2mM ATP	2mM ADP	2mM AMP	2mM GTP	2mM GDP	2mM NAD	2mM FAD	1% glycerol	5% glycerol	10% glycerol	20% glycerol	3% DMSO
G	1mM DTT	5mM DTT	10mM $CaCl_2$	10mM $MgCl_2$	10mM $MnCl_2$	10mM $ZnCl_2$	10mM $NiCl_2$	10mM $FeCl_3$	100mM KCl	100mM LiCl	200mM KSCN	10mM spermidine
H	10mM Arginine	10mM Alanine	10mM Methionine	100mM Glycine	3% Glucose	3% Galactose	10mM Urea	5% PEG400	0.5% n-Hexyl gluco-pyranosid	0.05% n-Decyl gluco-pyranosid	0.01% Triton X100	H_2O as reference

The second step was the actual buffer screen. The screen shown in Table 14 (row A-D) had been prepared in our laboratory and has been used for all conducted assays. Sufficient amounts of the protein

and the dye in the optimal concentrations were prepared and 40 µl buffer, 5 µl protein and 5 µl SYPRO Orange were mixed for all the conditions. The experiment was performed as described above and the optimal buffer condition, corresponding to the highest protein melting temperature, was selected.

An optional third step was an additive screen. The lower part of the screen in Table 14 (row F-H) contained different additives that might have stabilized a protein in solution. A sufficient amount of the eight times concentrated buffer, identified in the step before, the protein and the dye in the optimal concentrations were prepared and 5 µl buffer, 5 µl protein, 5 µl SYPRO Orange and 35 µl additive were mixed for all the conditions. The experiment was performed as described above and stabilizing or destabilizing additives were identified by their influence on the protein melting temperature.

3.2.4.3 Protein interaction studies

3.2.4.3.1 Pulldowns

High affinity interactions can be identified by the copurification or protein pulldown of the two binding partners. Ideally the two proteins carry different affinity tags, but also one tagged and one untagged protein can be used. To identify an interaction, the first protein is bound to the according affinity beads. The beads are washed and the second protein, which is not interacting with the affinity-beads, but possibly with the first protein bound to the beads, is added. The beads are washed again and are analyzed on an SDS-gel. If the second protein is present, it has been pulled down by the first protein, confirming an interaction of the two. For this experiment the two proteins can either be provided as lysates or as purified proteins. An unspecific interaction of the second protein with the beads can be excluded, by running a negative control without the first protein in parallel. The interaction can be verified by binding the second protein to the according affinity beads and adding the first protein afterwards. Usually 20 µl of affinity beads were incubated with either 1 ml of lysate, containing the first protein, or with 20 µg of the purified first protein in 1 ml of buffer, at 4°C for 30 min for protein binding. The beads were washed twice with 1 ml of washing buffer, before the second protein was added in 1 ml of lysate or buffer. The beads were incubated again at 4°C for 30 min, before they were washed twice with 1 ml of washing buffer, boiled in 20 µl SDS loading buffer and analyzed on an SDS-gel.

3.2.4.3.2 Size exclusion chromatography (SEC)

SEC can also be used to identify high affinity interactions. The two potential binding partners are mixed and separated on a SEC column. The elution is fractionated and analyzed on an SDS-gel. As a control, the two individual proteins are run on the same column. The two proteins are interacting, if in the mixture, compared to the two single proteins, a new absorption peak emerges at a lower elution volume. This new peak corresponds to the complex, which has a higher MW than the two individual proteins. The interaction can be verified by the presence of the two protein bands in the corresponding elution fractions on the SDS-gel. Usually 1 mg of each of the two proteins were mixed in total volume of 0.5 ml and incubated on ice for 30 min, before they were separated on a SEC column.

3.2.4.3.3 Isothermal titration calorimetry (ITC)

ITC is a universal method to identify protein-protein, but also protein-small compound and other interactions. It is possible to determine the binding affinity K_D, the stoichiometry n and the thermodynamic parameters ΔH and ΔS. An isothermal titration calorimeter contains two chambers, a reference cell filled with water or buffer and a sample cell filled with a solution of the first protein. Initially both cells are heated with a small and constant power. Thermoelements measure the temperature in both cells and a feedback element modifies the heating power of the sample cell, so that it is heated the same way as the reference cell. In the experiment the second protein or binding partner is repeatedly injected into the sample cell by a syringe and binds to the first protein in an endo- or exothermic reaction. The feedback element increases or decreases the heating power accordingly, to adjust the temperature in the sample cell to the one in the reference cell. In case of an exothermic reaction, less heating power

is needed to keep the temperature in both cells at the same level, whereas in the case of an endothermic reaction, the heating power needs to be increased. The time course of the electric heating power is recorded, resulting in a number of peaks, corresponding to the power adjustments during the individual injections. The area under the peaks corresponds to the thermodynamic heat that is released or absorbed in the reaction and from the course of the whole experiment the above mentioned parameters can be determined.

To prepare the experiment the two protein samples were dialyzed against 4 l of protein buffer at 4°C overnight to assure that the buffers were identical. If the binding affinity K_D was not known, the concentration of the protein in the cell was adjusted to 20 µM and the concentration of the protein in the syringe was adjusted to 200 µM. If the binding affinity K_D was already known from previous experiments, appropriate concentrations for the proteins were calculated, using Equation 1 and Equation 2.

$$\frac{c_{cell}}{K_D} \approx 20 - 100 \qquad \text{Equation 1}$$

$$\frac{c_{syringe}}{c_{cell}} \approx 10 \qquad \text{Equation 2}$$

The protein solutions were temperated to 23°C and degassed. Meanwhile the sample cell and the syringe were washed with water and buffer. The two proteins were filled into the sample cell and the syringe and the experiment was started. Normally the measurement was performed at 25°C and the protein from the syringe was added in 29 10 µl injections. Binding was recognized by the typical curve, showing high peaks in the beginning and decreasing peaks later in the experiment. Random peaks with amplitudes lower than 0.1 µcal/s indicated that no binding was present. In this case the experiment was repeated at least once, using a different temperature. Data analysis was performed, using the software package accompanying the ITC machine.

3.2.4.3.4 Biolayer interferometry (BLI)

Biolayer interferometry is a technique to quantify the interaction of two biomolecules, using an optical interference technique. An incident beam of white light is reflected on a reference surface and on a layer of biomolecules. Thereby a wavelength dependent interference pattern is produced. The presence of a binding partner on the surface causes an increase of the thickness of the biolayer and thereby a shift of the interference pattern by a certain wavelength $\Delta\lambda$ that is monitored in realtime. Usually one binding partner is immobilized on the surface, while the second binding partner is flushed over it and the interaction of the two is monitored as the time dependent wavelength shift.

BLI can be used to determine the binding affinity K_D and the kinetic parameters k_{on} and k_{off}, which are the association and dissociation rate constants. k_{on} and k_{off} are directly determined from the time dependent wavelength shift and K_D is calculated as the quotient of k_{off} and k_{on}. Another possibility to determine K_D is to use different concentrations of the binding partner and to calculate it from the equilibrium concentrations (or wavelength shifts) of the formed complex, using a Michaelis-Menten like model.

The BLI machine used in this work utilized optical fibers with functionalized tips. Usually tips coupled to an anti-His antibody were used to immobilize a His-tagged protein and a GST-tagged binding partner was tested for interaction. The machine allowed the parallel measurement of up to eight reactions. After coupling the His-tagged protein (at a concentration of 20 µg/ml) to the eight tips, the association and dissociation of the binding partner at seven different concentrations and one buffer blank were monitored. The tips were cleaned in BLI regeneration buffer before another seven concentrations were

assessed. The buffer composition as well as the concentration range of the binding partner had to be optimized for each experiment, here concentrations between 0.1 and 10 µg/ml were used. For this setup, the unspecific interaction of GST with the tips or the His-tagged protein had to be excluded in a control experiment. The binding affinity K_D was calculated from the equilibrium concentrations of the protein complex, using the accompanying software package.

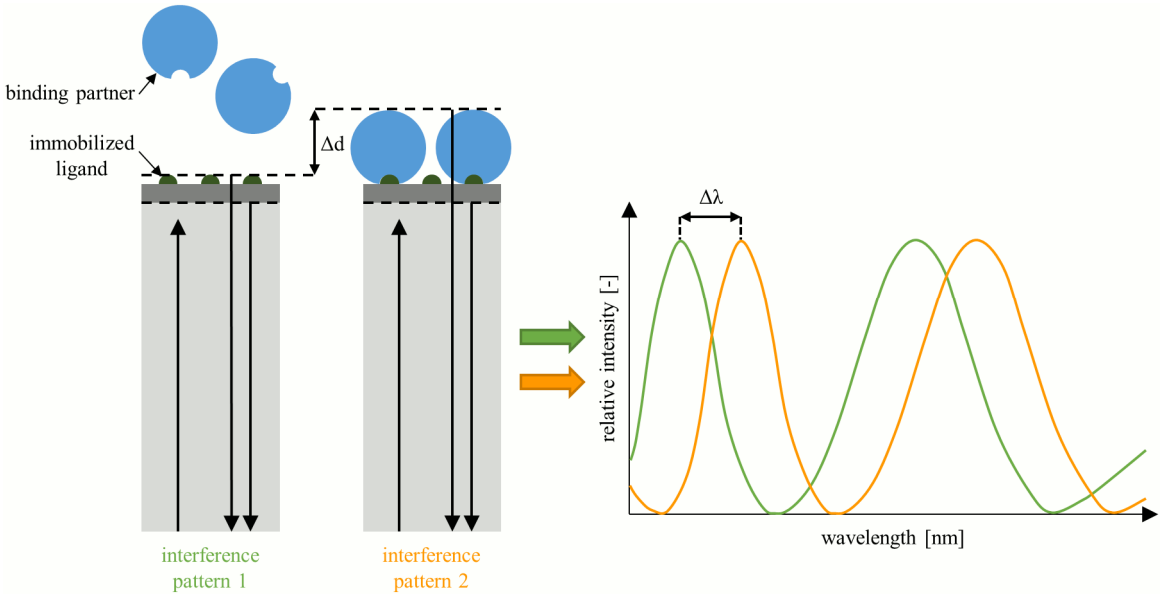


Figure 17: BLI setup to detect the binding of one protein to an immobilized ligand. The reflection of an incident beam of white light on two surfaces, produces a wavelength dependent interference pattern. One surface is an internal reference, the second surface is the layer of biomolecules. Interaction of the binding partner and the immobilized ligand increases the thickness of the layer and causes a shift of the interference pattern by a certain wavelength Δλ that is monitored in realtime.

3.2.4.3.5 Microscale thermophoresis (MST)

Thermophoresis is the property of small compounds to move in a temperature gradient, similar to electrophoresis, which is the property of charged particles to move in an electric field. The movement is dependent on the size, the charge and the hydration shell of the particles, which is represented by the Soret coefficient. For proteins, the movement is also influenced by conformational changes and the binding of other proteins or small compounds. Even binding a single ion influences the hydration shell and the Soret coefficient and thus the mobility of the protein. Microscale thermophoresis can be used to determine the binding affinity K_D of two proteins. For this purpose, one of the binding partners is labeled fluorescently. The concentration of the labeled binding partner has to be optimized, to exclude the saturation of the fluorescence detector, and is kept constant in the course of the experiment. It is mixed with a dilution series of the unlabeled binding partner and filled into capillaries. The buffer and the type of capillary have to be optimized in the first steps of an experiment, to exclude aggregate formation and the interaction of the proteins with the capillary surface. An infrared laser generates a temperature gradient of some Kelvin on a distance of a few micrometers, resulting in a thermophoretic force, experienced by the molecules. The thermophoretic force results in the movement of the proteins along the temperature gradient. The movement of labeled protein is monitored, by observing its fluorescence, using the same light path as the infrared laser. At higher concentrations of the binding partner, a complex is formed, which moves differently in the temperature gradient, due to its changed Soret coefficient, resulting in different equilibrium fluorescence values. Complex formation can be recognized by sigmoidal binding curves, if the normalized equilibrium fluorescence is plotted against the logarithmic concentration of the unlabeled binding partner. The binding affinity K_D is calculated by the accompanying software, using a Michaelis-Menten like model.

The MST machine used in this work measured 16 capillaries, corresponding to 16 concentrations of unlabeled protein, at once. One of the binding partners was fluorescently labeled, according to the supplied manual. Unlabeled protein concentration between about 10 nM to 100 µM were used.

3.2.4.4 Circular dichroism (CD) spectroscopy

Circular dichroism is the property of chiral compounds to differentially absorb left- and right handed circularly polarized light. In proteins not only the amino acids, but also the secondary structure elements have chiral properties and generate distinct CD-spectral signatures. For this reason CD-spectroscopy can be used to determine a proteins secondary structure composition.

A protein sample was dialyzed twice against 4 l of CD buffer, the first time for 3 h, the second time overnight. Afterwards its concentration was adjusted to 1 mM amide bonds. A cuvette with 1 mm light path was rinsed extensively with water and buffer. First a buffer blank and afterwards the protein sample was filled into the cuvette and the CD-signal was measured between 260 and 190 nm at 20°C. The CD-signal of the buffer blank was subtracted from the CD-signal of the sample, to obtain the CD-signal of the pure protein. Using this data, the accompanying software estimated the secondary structure composition of the protein.

3.2.4.5 Mass spectrometry (MS)

The following mass spectrometry analyses were performed by an in house service. For the identification of protein bands from a Coomassie stained SDS-gel, the bands were excised, washed three times with H_2O and dried in a vacuum concentrator for 10 min. The protein in the band was subjected to tryptic digestion and the sizes of the resulting peptides were determined by mass spectrometry. The protein was identified by searching a database, containing theoretical digestion patterns.

Edman degradation was used to identify the N-termini of protein fragments, resulting from limited proteolysis. An SDS-gel containing the fragment was blotted onto a PVDF-membrane. The membrane was stained with Coomassie and the fragment band was excised. It was washed three times with H_2O and dried in a vacuum concentrator for 10 min. The N-terminal residue of the fragment was repeatedly removed by the reaction with a phenylisothiocyanate, which was then identified chromatographically.

The MWs of peptides in a complex mixture of protein fragments, e.g. resulting from limited proteolysis, were identified by mass spectrometry and used to calculate their C-termini, using the known N-termini from Edman degradation. To perform this experiment a 30 µl sample was dialyzed against 4 l H_2O at 4°C for 2 h and given directly for the identification by mass spectrometry.

3.2.4.6 Small angle X-ray scattering (SAXS)

Small angle X-ray scattering is a method to obtain structural informations from a protein in solution. In contrast to X-ray crystallography no crystalline material is necessary. An incident X-ray beam falls onto a capillary, flushed with a protein solution. The X-rays are scattered and fall on a detector, where their distribution is monitored. Proteins in solution are randomly oriented, thus X-ray scattering results in radially decreasing intensity distributions (Figure 18). From the form of this intensity distribution, certain properties of the protein, like the average particle size and shape, can be deduced. The lower resolution limit of SAXS is about 10 Å, hence informations about the quaternary structure of a complex or the tertiary structure of a protein can be obtained. As not only the electrons in the protein, but also the electrons in the buffer contribute to the X-ray scattering, a buffer blank has to be measured and subtracted to obtain the scattering curve of the pure protein.

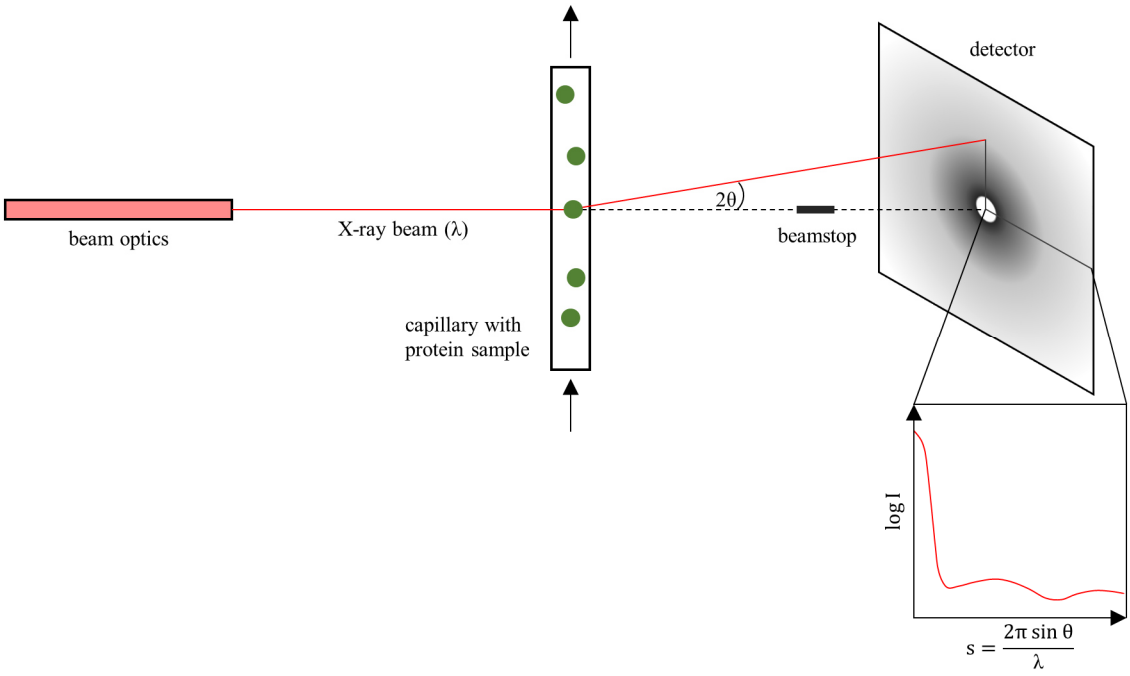


Figure 18: Setup of a typical SAXS experiment. An X-ray beam with wavelength λ is scattered by a protein solution in a capillary. The scattered X-rays fall on a detector. Because the proteins are randomly oriented, a radially decreasing scattering intensity distribution is observed. Normally the logarithmic intensity is plotted against the momentum transfer s. From the form of this scattering curve, certain properties of the protein can be deduced.

Using SAXS data, it is possible to create low resolution models of the protein. The problem is that a three dimensional model is calculated from a one dimensional scattering curve. The assignment is mathematically not unique, meaning that differently formed particles can produce the same scattering curve. For this reason a number of different models of the protein are created, aligned and averaged to obtain the most reliable features. SAXS can also be used to assess conformational changes of a protein or the arrangement of proteins in a complex by modifying the quaternary arrangement of different domains or proteins to fit the experimental data. It is also possible to determine the oligomerization state of a protein and to distinguish between several possible oligomers.

The ATSAS data analysis software is a program package, developed by the Biological Small Angle Scattering group at the EMBL in Hamburg Germany [169]. It contains a number of programs to work with and analyze small angle scattering data. Primus is a program to visualize and work with raw scattering curves, including operations like subtraction, averaging and merging of curves [170]. Crysol calculates the theoretical scattering curve of a protein with known structure and fits it to an experimental scattering curve [171]. A chi-value indicates the quality of the fit, the lower the chi-value, the better the fit. It can be used, if a high resolution model of the protein in solution is available and has to be verified. Gnom calculates the distance distribution function from the scattering curve [172], which is further used by dammif, a program that calculates the shape of a protein, using a bead model, which is optimized by simulated annealing [173]. The four programs damsel, damsup, damaver and damfilt are used to align and average the output models of several dammif runs [174]. Finally sasref is a program that modifies the quaternary arrangement of several subdomains or proteins, to fit an experimental SAXS scattering curve [175]. Also a number of other programs are included in the package that serve different functions. Using the online tool SAXS MoW, the MW of a protein can be determined from its SAXS scattering data [176].

To perform a SAXS experiment, a protein sample was dialyzed against 2 l of the protein buffer at 4°C overnight. It was centrifuged at 4°C, 13000 rpm for 30 min to remove particles and a dilution series was prepared, comprising 100 µl samples at concentrations of 5, 2.5, 1.25, 0.63 and 0.31 mg/ml. All

concentrations were measured spectrophotometrically. A large amount of buffer was filtered through a 100 kDa MW cutoff (MWCO) Vivaspin concentrator to remove particles and sent together with the protein sample to be measured. SAXS measurements were performed at BM29 of the ESRF in Grenoble, France and beamline P12 of PetraIII in Hamburg, Germany. SAXS data analysis was performed, using the ATSAS package.

3.2.5 Protein crystallization

X-ray crystallography is a method to obtain atomic resolution models of proteins. Other methods comprise NMR and cryo EM. However, with NMR it is difficult to work with proteins larger than 40 kDa and up to date there are only a limited number of atomic or near-atomic resolution structures from cryo EM. This makes X-ray crystallography the standard method in the structure determination of proteins. The bottleneck in X-ray crystallography is the production of protein crystals, which for some proteins is difficult and time consuming or even impossible.

To crystallize a protein, a solution of highly concentrated (generally 5-20 mg/ml) and pure (>95%) protein is necessary. A precipitant is added to decrease the solubility of the protein. This leads to supersaturation of the protein and in the best case to the formation of crystals. The phase diagram in Figure 19 shows the different phases, the protein solution adopts, depending on the concentrations of protein and precipitant. At low concentrations the phase is stable, the protein is soluble and no crystal formation is taking place. At intermediate concentrations a metastable phase is formed, which is bipartite. In the heterogeneous nucleation phase, crystal growth occurs only from preexisting nuclei, but no new nuclei are forming, while in the homogenous nucleation phase, spontaneous nucleation and crystal growth is taking place. At too high concentrations the phase is unstable and a precipitate forms, which is composed of solid state, non-crystalline protein. Thus for the formation of protein crystals, the concentrations of protein and precipitant have to be in the right range.

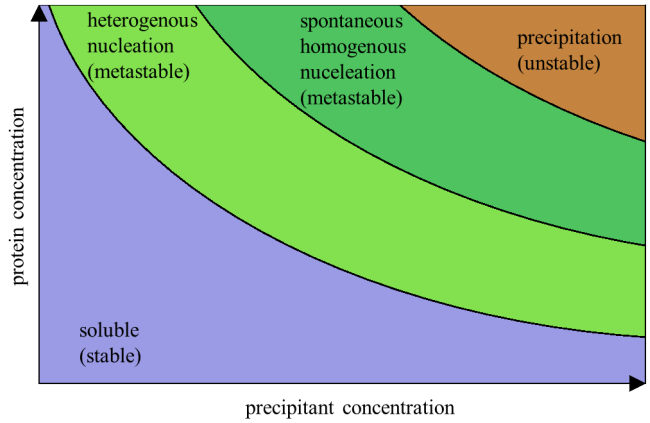


Figure 19: Crystallization phase diagram. At low protein and precipitant concentrations the phase is stable and no crystal formation is taking place. In the metastable heterogeneous nucleation phase, preexisting crystal nuclei can grow, but no new nuclei are forming. In the metastable spontaneous homogenous nucleation phase also new crystal nuclei are forming. Finally in the unstable precipitation phase the protein spontaneously precipitates, thus forming solid state, non-crystalline material.

Typical precipitants are salts, alcohols and polymers. Aside from the precipitant, also the pH and the presence of other chemicals, so called additives, can influence the crystallization of a protein. However, it is not possible to predict, in which combination and concentration ranges, crystal formation takes place. For this reason crystallization screens have been developed, to identify the crystallization conditions of a certain protein. In principal there are three different kinds of screens. Sparse matrix screens probe a wide range of different chemicals and crystallization conditions. In grid screens two variables like pH and precipitant concentration are systematically modified. Finally in random screens certain classes of chemical substances are mixed in random combinations and concentrations. A large number of screens is commercially available from different manufacturers, mostly sparse matrix and

grid screens. Random screens are usually used, when a first crystallization condition has been identified, to optimize this condition.

There are many different ways to set up a crystallization experiment, two of which are described here. In the sitting drop method a large reservoir is filled with a buffer of a certain composition (crystallization or reservoir solution). Apart from the reservoir, a small drop of highly concentrated protein is mixed with the buffer in a 1:1 ratio. The space, containing both the drop and the reservoir, is sealed (Figure 20). As the buffer, containing the protein, is 50% diluted, a net transfer of water from the drop to the reservoir occurs by vapor diffusion, until the chemical equilibrium is reestablished. In this process the drop volume shrinks and the concentrations of protein and precipitant rise. In the optimal case the solution is in the stable phase in the beginning and moves into the metastable phase in the course of the experiment, where first crystal nuclei form that grow into crystals. The hanging drop method is essentially similar, except that the protein drop is hanging under a cover slip (Figure 20).

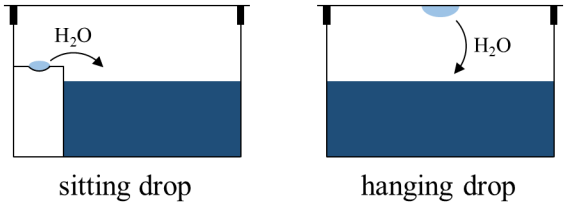


Figure 20: Crystallization experiment setups. In the sitting drop setup, a reservoir is filled with buffer of a certain composition. A protein drop is mixed with buffer in a 1:1 ratio and placed apart from the reservoir. The space is sealed to prevent drying out. Water is transferred by vapor diffusion from the drop to the reservoir, until an equilibrium is reached. The drop volume shrinks and the protein and precipitant concentrations rise. The hanging drop setup is similar, the difference is that the drop hangs under the coverslip that seals the space.

Initial crystallization screens were prepared, using a robotic system (96-well sitting drop setups). Usually protein concentrations of 5 and 10 mg/ml and a buffer control were used in initial screens. The reservoir contained 60-70 µl and the drop had a size of 0.4 µl (0.2 µl protein + 0.2 µl buffer). The plate was sealed, using Crystal Clear sealing tape, then it was incubated and monitored at 20°C or 4°C in plate incubators and regularly checked for initial crystallization conditions. These conditions were optimized in later steps, using self-prepared grid and random screens (24-well hanging drop setups). Here the reservoir contained 500-1000 µl and the drop had a size of 2 µl (1 µl protein + 1 µl buffer). The wells were sealed, using grease and cover slips. The aim was to obtain large three dimensional crystals, suitable for X-ray diffraction experiments.

To perform X-ray diffraction experiments, protein crystals have to be frozen in liquid nitrogen to avoid radiation damage. A cryo protectant solution is needed to prevent the formation of ice crystals during the freezing process. Normally the crystallization condition is supplemented with a cryo protectant, e.g. glycerol, and tested for the presence of water crystals in an X-ray diffraction experiment. The solution without ice crystals, containing the lowest cryo protectant concentration, is then used to freeze the protein crystals. The crystal is washed in the cryo protectant solution, mounted onto a nylon loop and shock frozen in liquid nitrogen. In this work, usually up to 25% of glycerol were used as cryo protectant. The frozen crystals were tested in house for diffraction and sent for data collection to a synchrotron, if suitable. Data was collected at PETRA III in Hamburg, Germany, at BESSY II in Berlin, Germany, at the SLS in Villigen, Switzerland and at the ESRF in Grenoble, France.

The setup of a typical X-ray diffraction experiment is shown in Figure 21. The crystal in its nylon loop is mounted on a goniometer head, which can be turned in several space dimensions. The crystal is cooled in place, by a nitrogen gas cryo stream. An incoming X-ray beam hits the crystal and is partially diffracted. The diffracted X-rays hit the detector, which is placed behind the crystal, while the undiffracted photons are blocked by the beamstop to prevent detector damage. The crystal is rotated and

a series of images is taken by the detector, recording the diffraction patterns. First the space group of the crystal is determined, by collecting four diffraction images at 90° angular distance. In this work, these images were indexed by iMosflm and a strategy was calculated for optimal data collection.

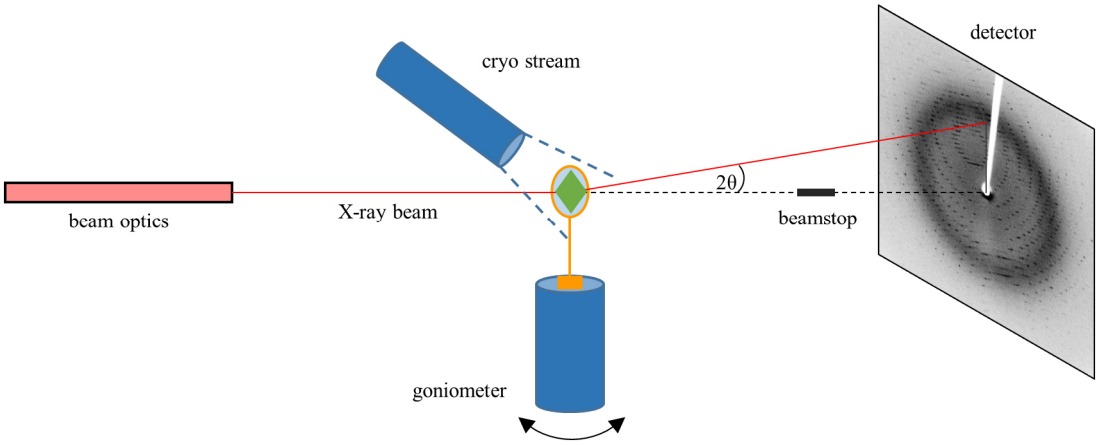


Figure 21: X-ray diffraction experiment setup. The crystal is mounted on a goniometer and cooled in place by a cryo stream. The incoming X-ray beam is partially diffracted and the detector records the diffraction images, while the crystal is rotated. The undiffracted X-rays hit the beamstop to prevent detector damage.

3.2.6 Structure Determination

A crystal is a macroscopic object that is composed of a translationally repeated general motif, the unit cell. The translational symmetry of a crystal is described by the crystal lattice, which can be thought of as a pattern of points in space. 14 Bravais lattices, grouped in 7 lattice systems, characterize the different possible translational symmetries. The three dimensional unit cell is defined by the length of the three lattice vectors a, b and c and by the three angles, enclosed by these vectors α, β and γ. 11 different point groups describe the possible internal unit cell symmetries of a protein crystal. The asymmetric unit is the smallest motif, by which the content of the unit cell can be reconstructed, using the point group symmetry operations (Figure 22 a). The combination of the 11 point groups with the 14 Bravais lattices results in 65 space groups, describing the possible symmetries of a protein crystal. Lattice planes are mathematical constructs that slice the crystal lattice in a distinct and periodic way. Sets of lattice planes are described by Miller indices (h k l) (Figure 22 b).

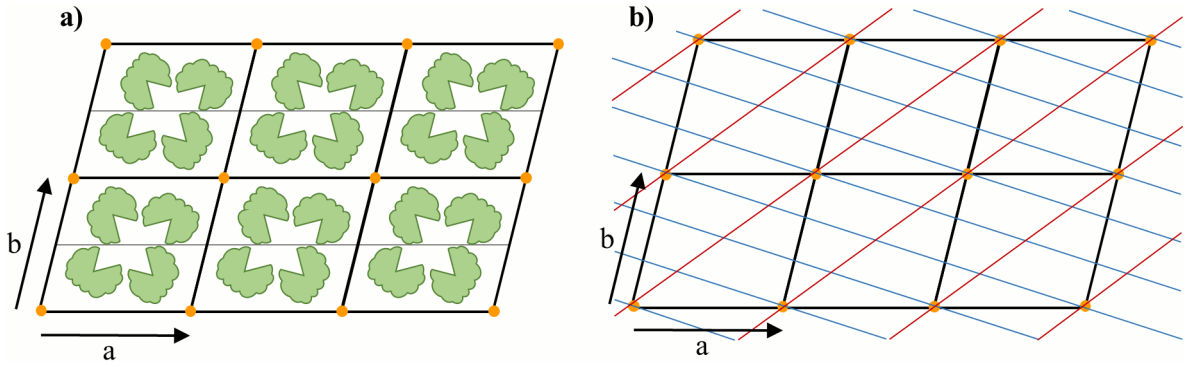


Figure 22: A two dimensional crystal. a) A crystal is composed of translationally repeated unit cells (black frames), characterized by the length of the unit cell vectors a and b. The translational symmetry is described by the crystal lattice (yellow dots). Here the unit cell has an internal C2 point group symmetry, thus the content of a unit cell is generated from the content of the asymmetric unit (grey line, half of the unit cell) by a 180° rotation. The asymmetric unit contains two copies of a protein, the unit cell four copies. b) Lattice planes slice the crystal lattice in a periodic way. Here two sets of lattice planes (blue/red lines) with Miller indices (1 3) and (1 -1) are shown.

The diffraction pattern of a crystal is characterized by distinct spots or reflections of different intensities, corresponding to constructively interfering diffracted X-rays (Figure 21). Bragg's law describes the

geometric requirements, which are necessary for constructive interference to occur (Equation 3). The diffraction of a crystal is interpreted as the reflection of X-rays on sets of lattice planes. Positive interference occurs if the path difference of the reflected waves ($2d_{hkl}\sin\theta$) is an integer multiple of the wavelength ($n\lambda$) (Figure 23).

$$n\lambda = 2d_{hkl}\sin\theta \qquad \text{Equation 3}$$

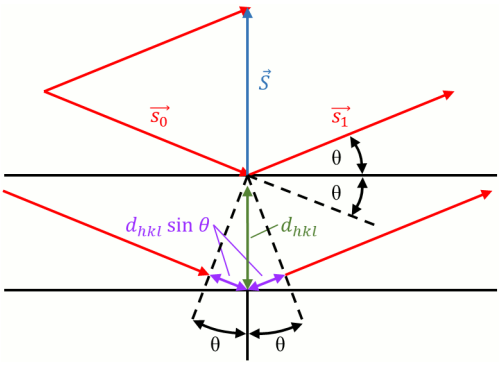


Figure 23: Diffraction of X-rays in a crystal is interpreted as the reflection of X-rays on sets of lattice planes. $\overrightarrow{s_0}$ and $\overrightarrow{s_1}$ are the incoming and diffracted wave vectors, $\vec{S}$ is the scattering vector ($\vec{S} = \overrightarrow{s_1} - \overrightarrow{s_0}$), θ is the scattering angle of the reflected wave and d_{hkl} is the interplanar distance of a set of lattice planes (hkl) (modified from [177]).

Like lattice planes, the reciprocal lattice is a mathematical concept. It greatly simplifies the description of diffraction events and of the conditions, necessary for diffraction to occur. Reciprocal space is characterized by inverse relations to real space. Sets of lattice planes in real space are represented by reciprocal lattice vectors $\overrightarrow{d^*_{hkl}}$ in reciprocal space. Using the concept of reciprocal space, the diffraction condition from Bragg's law can be visualized by the Ewald construction. The scattering diagram from Figure 23 is placed inside an Ewald sphere of radius 1/λ (Figure 24). From the diagram the relation $|\vec{S}| = 2\sin\theta/\lambda$ can be derived. The origin of the reciprocal lattice is placed on the right end of the Ewald sphere, where the extended incoming X-ray beam would leave the sphere. For a reciprocal lattice point, lying on the surface of the Ewald sphere, the condition $\left|\overrightarrow{d^*_{hkl}}\right| = |\vec{S}| = 2\sin\theta/\lambda$ is true, which corresponds to the constructive interference condition of Bragg's law. Consequently reflections, corresponding to all the lattice planes, whose reciprocal lattice vectors lie on the surface of the Ewald sphere, can be observed.

The question remains, how to reconstruct the electron density in a crystal from the recorded diffraction patterns. To find an answer, first the diffraction of a single atom is assessed, followed by the more complicated diffraction of molecules and crystals. The atomic scattering factor f_S^0 is a complex function that describes the diffraction of a single atom (Equation 4).

$$f_S^0 = \int_{\vec{r}}^{V(atom)} \rho(\vec{r})\, e^{2\pi i \vec{S}\vec{r}}\, d\vec{r} \qquad \text{Equation 4}$$

$\rho(\vec{r})$ describes the electron density of the atom and $\vec{S}$ is the scattering vector. The formula has the form of a Fourier integral, hence the electron density of the atom is physically Fourier transformed, resulting in a bell shaped intensity distribution. The diffraction patterns of molecules and whole crystals result from the addition of the contributions of the structure forming atoms. Diffraction of a single molecule produces a continuous and irregular, but centrosymmetric intensity distribution. The diffraction pattern

of a molecular crystal shows discrete spots with varying intensities. It can be described as the scattering function of the molecule, which is amplified by the number of molecules in the crystal and sampled in discrete spots. The positions of these spots can be derived from Bragg's law, as described above.

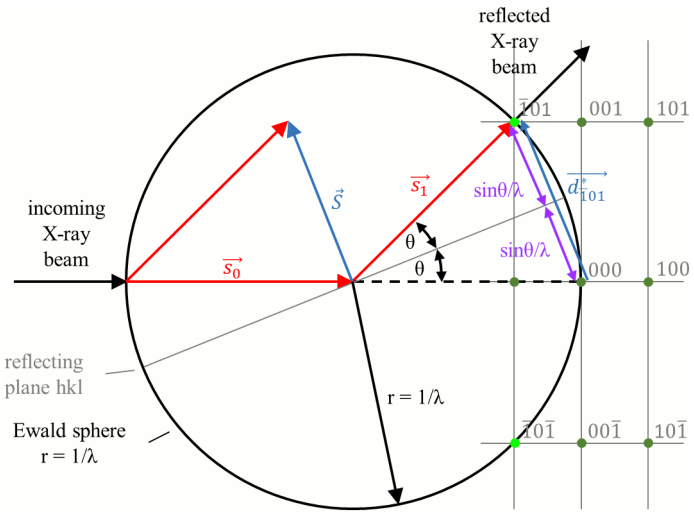


Figure 24: The scattering diagram from Figure 23 is placed inside a sphere of radius 1/λ, the Ewald sphere. The relation $|\vec{S}| = 2\sin\theta/\lambda$ can be derived from the diagram. The origin (000) of the reciprocal lattice is placed on the right end of the Ewald sphere. If a reciprocal lattice point, described by a vector $\overrightarrow{d^*_{hkl}}$, lies on the surface of the Ewald sphere (bright green lattice points), the condition $|\overrightarrow{d^*_{hkl}}| = |\vec{S}| = 2\sin\theta/\lambda$ is true, which corresponds to the constructive interference condition from Bragg's law. Consequently the corresponding reflections can be observed on the detector (modified from [177]).

The formula for the diffraction of a molecular crystal can be rearranged to calculate the complex structure factor F_h, which contains both the intensity and the phase information of a single reflection, described by the Miller indices of the corresponding set of lattice planes $\vec{h} = (hkl)$ (Equation 5).

$$F_h = \sum_{j=1}^{atoms} f^0_{S,j} \cdot e^{2\pi i \vec{h}\overrightarrow{x_j}} \qquad \text{Equation 5}$$

$f^0_{S,j}$ describes the atomic scattering factor and $\overrightarrow{x_j}$ the coordinates of atom j. The complex structure factor F_h contains contributions of all atoms in the unit cell. Now the electron density in the crystal $\rho(\vec{x})$ can be reconstructed from the complex structure factors, using inverse Fourier summation (Equation 6).

$$\rho(\vec{x}) = \frac{2}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F_h \cos\left(2\pi\vec{h}\vec{x} - \alpha_h\right) \qquad \text{Equation 6}$$

F_h are here the structure factor amplitudes and α_h the structure factor phases. $\vec{h} = (hkl)$ are the Miller indices of the corresponding sets of lattice planes. A problem arises, because the structure factor is complex, but only the intensity I of a reflection, which is proportional to $|F_h|^2$, but not the phase can be measured by the detector. Hence the phase information has to be provided from other sources, a situation known as the phase problem. One solution is to calculate the phases, using the known structure of a similar protein, and to use them to calculate the electron density of the new structure, a method called molecular replacement. If no suitable structure is available, experimental methods can be used to obtain the needed phase information. Multi wavelength anomalous dispersion (MAD) uses the anomalous dispersion of heavy metal ions, which have to be incorporated in the crystal, to obtain the phase information. Normally three datasets, using different X-ray wavelength are collected: first at the absorption peak of the anomalous scatterer, second at the inflection point of the absorption curve and

third at a remote wavelength. Heavy metal ions can be incorporated in a crystal by soaking into a preformed crystal, by cocrystallization or by producing and crystallizing selenomethionine labeled protein. Single wavelength anomalous dispersion (SAD) needs only one dataset, collected at the absorption peak wavelength of the anomalous scatterer, and uses density modification methods to improve the initial phases. Multiple isomorphous replacement (MIR) needs at least two anomalous datasets with different heavy metals, incorporated in the crystal, additional to the dataset of the native crystal. For single isomorphous replacement (SIR) only one anomalous dataset is necessary and density modification methods are used to improve the phases.

In this work, all necessary calculations for the structure determination were performed by computer programs. XDS was used to identify and index the spots in the images, then to integrate and scale their intensities [178]. In the XDS output files the most probable space group was indicated, which was checked, using Pointless. Pointless also tested for the presence of screw axes and checked for twinning [179]. The data was reprocessed, using XDS, which was provided with the correct space group. XDS included a scaling function, but XSCALE was used to rescale the data, utilizing provided resolution shells [178]. XDSCONV was used to convert the XSCALE ahkl output file to a Phenix mtz input file [178]. It also marked 5% of the data with an R-free flag, meaning this part of the data was not used in later refinement, but in the validation of the structure. The Phenix program Phaser-MR was used for molecular replacement [180], [181]. Phenix AutoBuild was used for automated model building [182], while WinCoot [183] and phenix.refine [184] were used for iterative manual model building and refinement steps.

Several R-values are used for the evaluation of the quality of the data and the model. R_{meas} describes the quality of the data, or to what extent symmetry related observations of the same reflection differ from each other (Equation 7).

$$R_{meas} = \frac{\sum_{hkl} \sqrt{\frac{n}{n-1}} \sum_{j=1}^{n} \left| I_{hkl,j} - \langle I_{hkl} \rangle \right|}{\sum_{hkl} \sum_{j} I_{hkl,j}} \qquad \text{Equation 7}$$

R_{work} measures how good the refined model describes the experimental data, using a working dataset. R_{free} is calculated the same way as R_{work}, but a separate data set of randomly selected reflections is used, which are never used in refinement (Equation 8). R_{free} is always larger than R_{work}, but the difference should not rise above 5%, as this indicates overfitting of the model.

$$R_{work/free} = \frac{\sum_{hkl} \left| F_{hkl}^{obs} - F_{hkl}^{calc} \right|}{\sum_{hkl} F_{hkl}^{obs}} \qquad \text{Equation 8}$$

3.2.7 The ESPRIT system

The ESPRIT system has been developed in the laboratories of Darren Hart and is located at the EMBL in Grenoble, France. It is a system to screen for soluble truncated constructs of difficult to express proteins. This is accomplished by the generation of a truncation library of the gene of interest (GOI) on the DNA level and by the subsequent screening of a large number of clones for soluble protein expression. The GOI can be truncated from the N-terminus, the C-terminus or from both sides. The soluble protein expression is quantified by a colony blotting technique, detecting N- and C-terminal signals with fluorescent probes. The 96 most promising clones are further analyzed by small scale test-expressions and the protein quantity is assessed by Western blotting and Coomassie stained SDS-gels. The process is highly automated by a robotic system for picking and handling of clones and a liquid handling system for test-expressions [185].

The GOI has to be cloned into the pESPRIT002 vector, using the *Asc*I and *Nsi*I restriction sites. This vector contains an N-terminal His-tag and a C-terminal biotin acceptor peptide (BAP). For the detection of soluble protein, the His-tag can later be detected by fluorescently labeled antibodies. The BAP is coupled to biotin by *E. coli*, which can be detected by a fluorescently labeled streptavidin (Figure 25). If a second protein is to be coexpressed, it is cloned into the pACYC vector, using *Bam*HI and *Not*I restriction sites. This vector contains an N-terminal Flag-tag. The cloning steps are performed at the home laboratory.

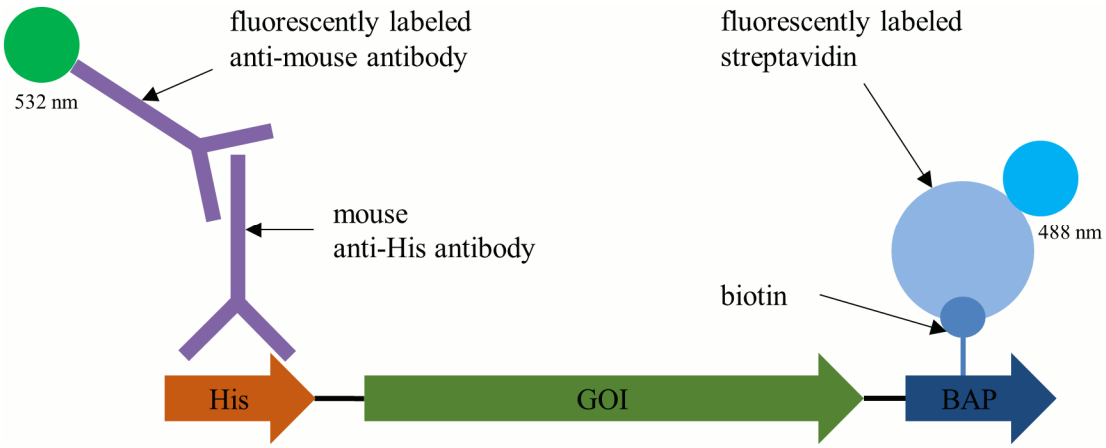


Figure 25: The GOI is cloned into the pESPRIT002 vector that contains an N-terminal His-tag and a C-terminal BAP. Soluble expression of the resulting protein can be assessed by the detection of the His-tag, using fluorescently labeled antibodies, and the detection of the BAP, by a fluorescently labeled streptavidin.

The next step is the generation of the truncation library, which is performed by the technical staff of the ESPRIT system. For the generation of an N-terminal truncation library of the GOI, the vector is cut open, using *Aat*II and *Asc*I restriction enzymes. *Asc*I generates a 5'-overhang, which is sensitive to exonuclease III degradation, while *Aat*II generates a 3'-overhang, which is not. By using this strategy, it can be assured that exonuclease III only digests the GOI, but not the backbone of the vector. After opening the vector, exonuclease III is added and the solution is incubated at 22°C for 1 h. 1/60th of the volume is taken every minute and put into a quenching tube, containing 3 M NaCl. After 1 h the exonuclease is denatured, by heating the quenching tube to 70°C for 20 min. The following treatment with mung bean nuclease at 30°C for 30 min and a *Pfu* polishing step at 72°C for 20 min generates blunt ends. The DNA is separated on an agarose gel and the library is excised from the gel in the preferred size range. The library can also be separated into sub-libraries, representing different sized fragments, to prevent the domination of small clones in the screen. The DNA library is extracted from the gel, ligated and transformed into chemocompetent OmniMAX cells by heat shock. 0.5 ml of LB are added and the cells are incubated at 37°C, 180 rpm for 1 h. A part is plated on LB-agar plates and incubated at 37°C overnight. The rest of the cells is stored at 4°C until further use. The quality of the library is assessed either by screening 48 clones by colony PCR or by performing minipreps and restriction enzyme digestion of 20 clones. If the size range and distribution is acceptable, the remainder of the cells are plated on a Q-tray, containing LB-agar, and incubated at 37°C overnight. A Q-tray is a large agar plate of about 25 x 25 cm, containing 300 ml of LB-agar (compare Figure 26). The colonies are scratched from the plate, using PBS and a Drigalski spatula, made from a Pasteur pipette. The DNA is purified, using a QIAPrep Spin Miniprep Kit.

At this stage of the experiment, a three week stay is organized to work with Philippe Mas, the technical assistant responsible for running and maintaining the ESPRIT-system. In the next step 60 ng of the library were transformed into one aliquot of 40 µl electrocompetent BL21 AI cells and 1 ml of LB was added. The cells were incubated at 37°C, 180 rpm for 1 h, before they were plated on a Q-tray, which

was incubated at 37°C overnight, then at room temperature for another night. This way the colonies became whiter and less translucent, thus improving the contrast for their automatic recognition. Six to eight thousand colonies per Q-tray should have been obtained, otherwise the amount of BL21 cells had to be modified before plating. 384-well plates were filled with 80 µl of TB, supplemented with HMFM, per well. Colonies were picked from the Q-trays into the 384-well plates, using a robotic system (Figure 26). The number of picked colonies was chosen, so that every construct was represented three times at a one bp resolution. A maximum of about 28000 clones could be handled by the system. The 384-well plates were incubated at 37°C, 300 rpm overnight and then stored at -80°C.

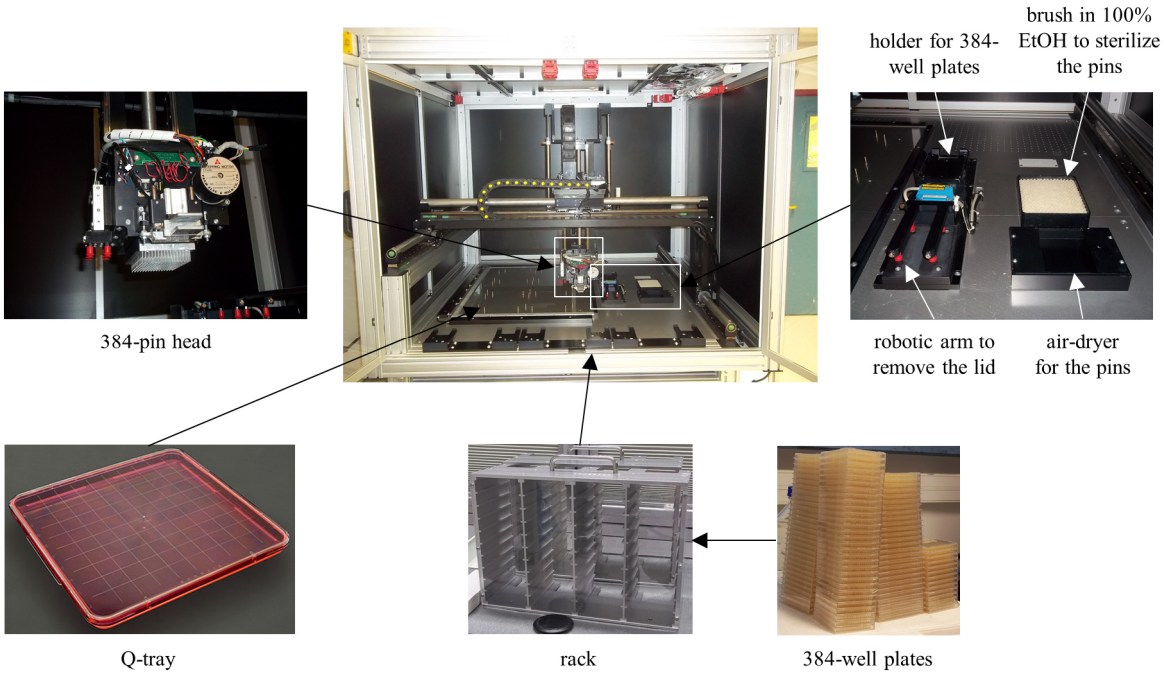


Figure 26: The Kbiosystems robotic solutions picking and gridding robot. A head contained 384 pins, which could be individually moved. The head was brushed in 100% EtOH, then air dried to sterilize it. A camera, attached to the head, scanned the Q-tray and a software recognized the colonies. 384 colonies were picked by the head and transferred to a 384-well plate, before the head was sterilized again and continued to pick further colonies. Up to four Q-trays and 120 384-well plates could be provided. The system was also able to blot 384-well plates on a nitrocellulose membrane and to inoculate one 384-well plate from a different one.

The frozen plates were replicated by the robotic system into fresh 384-well plates, filled with 80 µl TB, supplemented with TB salts, per well. The plates were incubated at 37°C, 300 rpm overnight. In parallel, a positive control, expressing MBP, was incubated in TB at 37°C, 180 rpm overnight. In every 8th 384-well plate, the clone in A1 was replaced by the positive control. A square shaped nitrocellulose membrane was put on a Q-Tray, containing LB-agar. The clones from up to 48 384-well plates were arrayed in duplicates onto the membrane by the robotic system, using a certain pattern shown in Figure 27, which had the advantage that the clones could be correctly assigned, even if the square shaped membrane was unintentionally rotated. The membrane on the Q-tray was incubated at 25°C overnight. Then it was transferred to a new Q-tray, containing LB-agar, supplemented with 50 µM biotin and 0.2% arabinose, to induce protein production and incubated at 30°C for 4 h.

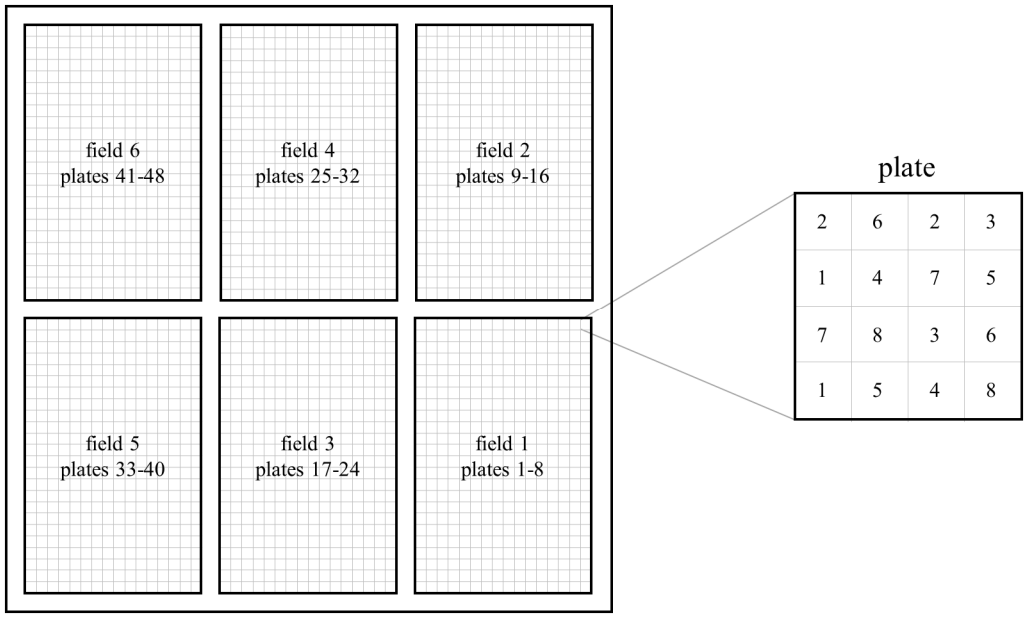


Figure 27: The clones of up to 48 384-well plates were blotted onto a square shaped nitrocellulose membrane. This membrane was divided into six fields, each containing the clones from eight 384-well plates. Each field contained 384 smaller 4x4 arrays of colonies, each array containing the clones of a certain well from all eight 384-well plates in duplicates. The clones in the 4x4 arrays had a certain layout, which had the advantage that they could be correctly assigned, even if the square shaped membrane was unintentionally rotated. The colonies in the 4x4 array had a distance of 0.8 mm. The scan of a real membrane is shown in Figure 36.

The membrane was transferred onto a Whatman paper, soaked in denaturation solution, for 10 min to lyse the cells. Then it was transferred onto a new Whatman paper, soaked in neutralization solution, for 5 min twice to return to a neutral pH. Finally the membrane was submersed in SSC (2x) for 10 min. Now the proteins, which were contained in the cells, were bound to the nitrocellulose membrane and the residual cell debris was removed by scraping with a Drigalski spatula, made from a Pasteur pipette. The membrane was blocked by adding PBS-T, supplemented with rabbit-serum, at 4°C for 1 h. Then it was washed with PBS-T for 5 min and the first antibody (mouse anti-His) was added in 50 ml PBS-T and incubated at 4°C for 1 h. The membrane was washed in PBS-T three times for 2 min, before the second antibody (fluorescently labeled anti-mouse) and a fluorescently labeled streptavidin (Streptavidin Alexa Fluor 488 conjugate, Invitrogen) were added in 50 ml PBS-T and incubated at 4°C for 1 h. The membrane was washed again in PBS-T three times for 2 min and then scanned, using a fluorescence scanner. The fluorescently labeled anti-mouse antibody had an emission wavelength of 532 nm, while the fluorescently labeled streptavidin had an emission wavelength of 488 nm. From here on, the fluorescence intensity values of the fluorescently labeled antibody is referred to as His-signal and the fluorescence intensity values of the labeled streptavidin as BAP-signal.

The fluorescence values were analyzed, using the VisualGrid software, which converted them into a table format that was further analyzed in Microsoft Excel. A histogram was generated that divided the His-signals into 15 groups. The one group, containing the highest number of clones, was defined as background and was excluded from further analysis, together with the groups containing lower intensities. From the rest of the clones only the fourth quartile with the highest His-signals was included in further analysis. These clones were sorted according to their BAP-signals. The positive controls were excluded and finally the 95 clones with the highest BAP-signals were selected for further analysis. One positive control was included as the 96th clone.

Statistically two third of the N-terminally truncated constructs are out of frame, resulting in the absence of the BAP and low BAP-signals. If insoluble protein is produced, the BAP is less accessible and thus not coupled to biotin efficiently, also resulting in low BAP-signals. Consequently only clones with both

high His- and BAP-signals should be highly soluble, in frame constructs, which are interesting for further analysis.

A 96-well deep well block was filled with 0.5 ml LB each and inoculated with the culture from the 384-well plates of the 96 best clones. The deep well block was closed, using Qiagen AirPore Tape Sheets and incubated at 37°C, 300 rpm overnight. 10% glycerol were added to each well, the deep well block was shaken again at 37°C, 300 rpm for 15 min and then frozen at -80°C.

These 96 most promising clones were further analyzed by test-expressions in the next step. A 96-well deep well block was filled with 0.5 ml TB each, inoculated from the frozen deep well block and incubated at 37°C, 300 rpm overnight as a preculture. 24-well deep well plates were filled with 4 ml TB each, inoculated with 50 µl of the preculture and grown to OD 0.8-1.0, before they were induced with 0.2% arabinose. Then they were incubated at 25°C, 300 rpm overnight. The cells were spun down at 4°C, 3700 rpm for 10 min. Each pellet was resuspended in 4 ml spheroplast formation buffer and incubated at 25°C, 300 rpm for 15 min, before the cells were spun down again at 4°C, 3700 rpm for 10 min. The buffer was removed carefully and the cells were lysed by freezing at -80°C for 30 min. Then the pellets were thawed at room temperature for some minutes, resuspended in 800 µl lysis buffer each and incubated at 25°C, 300 rpm for 5 min. The four 24-well plates were reformatted into a 96-well deep well block by a TECAN liquid handling system and the deep well block was centrifuged at 4°C, 3700 rpm for 30 min.

A 96-well TECAN receiver plate was loaded with 75 µl Ni-NTA beads each. The beads were washed with 700 µl H_2O and 700 µl washing buffer each. The supernatant was transferred from the deep well block to the receiver plate with the His-affinity beads and incubated at 4°C for 30 min for protein binding. The TECAN liquid handling system washed the beads three times with 300 µl washing buffer and then added 60 µl elution buffer each. The beads were incubated in elution buffer for some minutes, before the buffer was removed and collected in a 96-well PCR plate, by centrifugation at 4°C, 100 rcf for 5 min.

20 µl samples were loaded and run on SDS-gels, which were all blotted on the same membrane. The membrane was washed once with PBS-T, before it was blocked with 100 ml of blocking solution at 4°C overnight. The membrane was washed once with PBS-T for 5 min, before it was incubated with 50 ml of blocking solution, supplemented with the first antibody (rabbit anti-flag), at 4°C for 1 h. The membrane was washed with PBS-T three times for 5 min. The second fluorescently labeled antibody (Alexa Fluor 633 goat anti-rabbit IgG (H+L), Invitrogen) and fluorescently labeled streptavidin (Streptavidin Alexa Fluor 488 conjugate, Invitrogen) were added in 50 ml PBS-T and incubated at 4°C for 1 h. The membrane was washed again in PBS-T three times for 5 min and dried between two Whatman papers. Then the membrane was scanned in the fluorescence scanner, which detected the C-terminal BAP of the GOI and the N-terminal Flag-tag of a coexpressed and copurified binding partner. All or a number of constructs were selected and samples were analyzed on SDS-gels, which were Coomassie stained, to assess the quantity of protein expression. Finally the 96 clones were sent back to the home laboratory.

In the subsequent steps, the 96 clones were first spread on LB-agar plates and single colonies were picked and sequenced to determine the construct boundaries. The protein quality and quantity of interesting constructs was assessed by test-expressions and purifications. The TEV-cleavage of the His-tag was tested as well as the oligomerization state of the protein on SEC. The last step before large scale protein production was the removal of the BAP and addition of a stop codon, by *Bsp*EI digestion and religation of the vector.

4 Results

4.1 VirG

4.1.1 Testing of the initial VirG-constructs

The objective of this part of the work was the crystallization and the structural characterization of VirG, which required milligram amounts of soluble protein. For this reason, in a first step, constructs and expression conditions were determined that yielded sufficient amounts of protein, to perform the necessary crystallization experiments.

Initially eight different Strep-tagged VirG-constructs (VirGK1-K8) were designed, based on predicted and experimentally determined domain boundaries, as well as secondary structure prediction. The smallest of these constructs (VirGK5) comprised only the putative IcsB/ATG5 binding domain, the largest one (VirGK4) nearly the whole α-domain (Figure 28). Test-expressions were done in Rosetta 2 (DE3) cells, but only faint soluble expression for VirGK2, K3, K6, K7 and K8 was observed (Figure 29 a).

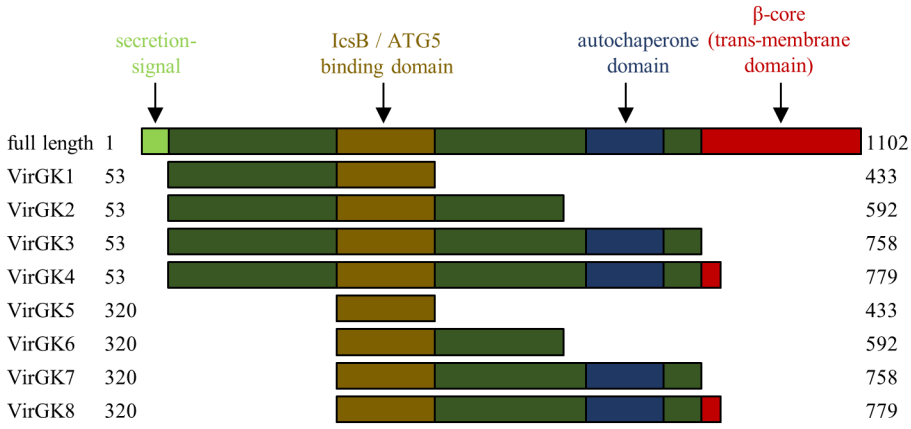


Figure 28: Overview of the eight initial VirG-constructs, VirGK1-K8.

One strategy to increase the soluble expression of a difficult to express protein is to fuse it to a solubility tag like GST, Sumo or MBP. For this reason the same constructs (VirGK1-K8) were cloned as GST-fusion proteins and test-expressions were performed (Figure 29 b). The GST-tagged constructs expressed slightly better than the Strep-tagged constructs, as soluble expression was observed for all of them, but still most protein bands were quite faint. GST-VirGK2, K3, K4 and K6 showed stronger expression than the rest of the constructs, but the amounts were still not sufficient.

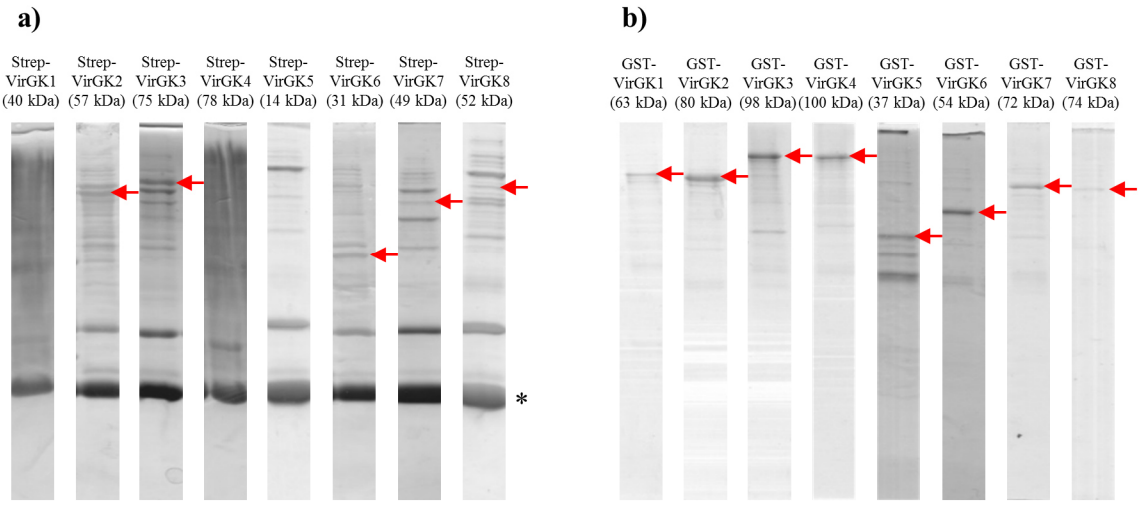


Figure 29: a) Coomassie stained SDS-gels for Strep-VirG-constructs test-expressions; b) Coomassie stained SDS-gels for GST-VirG-constructs test-expressions. Only soluble protein expression for the best expression condition is shown: TB, 20°C, 200 µM IPTG, overnight. Red arrows mark the corresponding VirG protein bands if present. The asterisk marks the streptavidin band, originating from the affinity beads. For the Strep-VirG-constructs only weak expression was observed for K2, K3 and K6-K8. For the GST-VirG-constructs expression for all constructs was observed, with stronger expression for K2-K4 and K6.

A different way to increase soluble protein yields, is coexpression with binding partners. In a complex, the solvent exposed protein surface is dramatically changed and the expression of the complex might be much better than the expression of the individual proteins. Hence, as a next step, the same Strep- and GST-tagged VirG-constructs were coexpressed with His- and Strep-tagged IcsB and ATG5 respectively (Figure 30 a, b). Pulldowns for both proteins in the complex were done, to check for the presence of both binding partners. IcsB and ATG5 were expressed strongly in all cases. For this reason, only the Strep- and GST-pulldowns, for the verification of the presence of the VirG-constructs, are shown here. Soluble protein expression levels of the VirG-constructs, using coexpression, were similar to VirG expression without coexpressed IcsB or ATG5. For both the Strep- and GST-fusion proteins, expression of nearly the same constructs was observed (compare Figure 29, Figure 30). Consequently the coexpression of the two binding partners IcsB and ATG5 didn't have a positive effect on the soluble expression of VirG. If VirG and IcsB/ATG5 interacted as expected, the copurification of an additional IcsB/ATG5 band, of similar strength like the VirG band, would have been expected. As the VirG expression was generally weak, it was not possible to verify these interactions in this experiment.

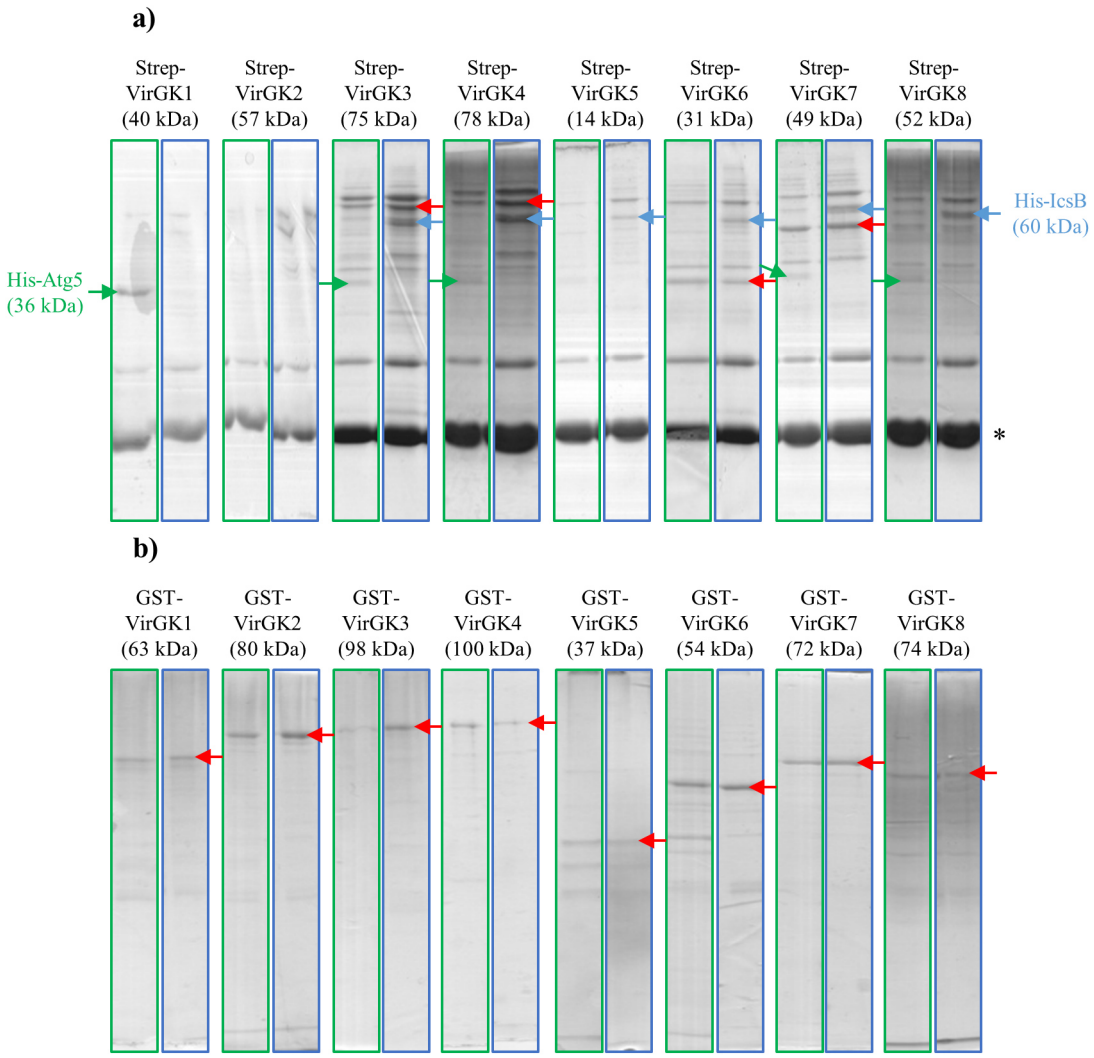


Figure 30: a) Coomassie stained SDS-gels for Strep-pulldowns of Strep-VirG, His-IcsB/ATG5 test-coexpressions. b) Coomassie stained SDS-gels for GST-pulldowns of GST-VirG, Strep-IcsB/ATG5 test-coexpressions. For each VirG-construct the left lane corresponds to coexpression with ATG5 (green frame), the right lane with IcsB (blue frame). Only soluble protein expression for the best expression condition is shown: TB, 20°C, 200 µM IPTG, overnight. Red arrows mark the corresponding VirG protein band if present, blue arrows IcsB and green arrows ATG5. The asterisk marks the streptavidin band, originating from the affinity beads. Strep-VirGK3, K4, K6 and K7 showed weak expression. All GST-VirG-constructs expressed solubly, with stronger expression for K2, K3, K6 and K7.

As VirG is an outer membrane protein and is normally secreted *via* the sec-pathway, in the next step periplasmic expression of four VirG-constructs (VirGK1-K4) was tested (Figure 31). For this purpose an N-terminal ompA secretion signal peptide [186] was cloned in front of the Strep-tag of the four constructs. Expression was observed for all four constructs, which was an improvement compared to the intracellular Strep-VirG expression, but the overall expression levels were similar to cytoplasmic GST-VirG expression levels. Hence, also periplasmic expression did not significantly increase the yields of soluble VirG protein.

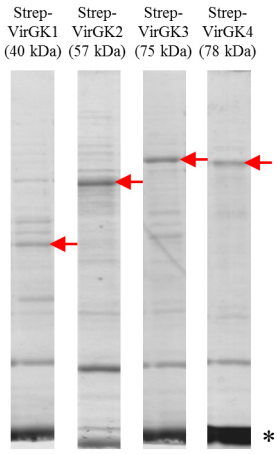


Figure 31: Coomassie stained SDS-gels for Strep-VirG-constructs periplasmic test-expressions. Only soluble protein expression for the best condition is shown: TB, 20°C, 200 µM IPTG, overnight. Red arrows mark the corresponding VirG protein band. The asterisk marks the streptavidin band, originating from the affinity beads. Expression was observed for all four constructs, with stronger expression for VirGK2 and K3.

For all the preceding approaches, only soluble protein expression levels are shown (Figure 29-Figure 31). The reason is that in all cases the expression levels in the insoluble fraction were low or not present. Hence not only the solubility, but the general expression of VirGK1-K8 was problematic. The reason might have been that translation was not taking place or that the translated protein was quickly degraded.

As further improvement of the VirG expression levels was not possible, 1 l large scale expressions and purifications were done for the best expressing GST-VirG fusion constructs (VirGK2, K3, K4 and K6), to increase the yield by increasing the volume. The elution fractions from the GST-affinity column were analyzed on an SDS-gel. A second, identical gel was blotted on a PVDF-membrane and the presence of the GST-fusion proteins was detected by Western blotting, using primary anti-GST antibodies (Figure 32). In all cases protein expression was observed, the strongest for VirGK6, but also a ladder of C-terminally truncated GST-tagged constructs was generated. Probably premature termination of translation was the reason for this phenomenon.

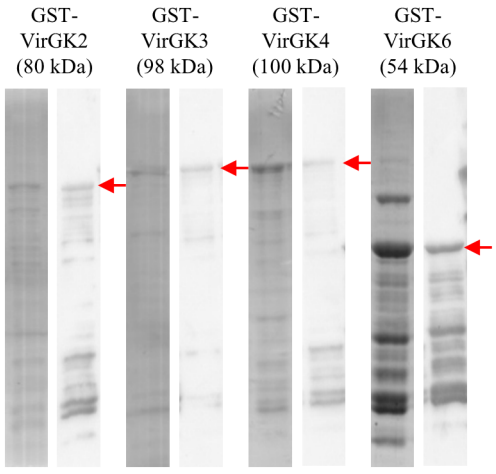


Figure 32: For each GST-VirG-construct in the left lane a Coomassie stained SDS-gel of the first elution fraction of the GST-affinity column and in the right lane the corresponding Western blot with anti-GST antibodies is shown. Red arrows mark the corresponding VirG protein band. Except for VirGK3, where the overall expression levels were weak, a ladder of C-terminally truncated constructs was seen, starting with individual GST at the lower MW end and ending with the full length GST-VirG fusion protein at the upper end.

The GST-tags of VirGK3 and K6 were removed by digestion with TEV-protease. The protein sizes before and after the digestion were compared on an SDS-gel. Only VirGK3 was digested but not VirGK6, as only for the former a shift in the height of the protein band was observed. The identity of

the proteins was confirmed by Western blotting before. Hence poor accessibility of the TEV protease cleavage site might be the reason for the inefficient cleavage in the case of VirGK6. After digestion, both proteins were separated on an S200 16/60 column. Both proteins eluted in the void volume of the column, indicating that they formed soluble aggregates. This was problematic, as aggregated protein is usually not suitable for crystallization.

As VirGK3 was digested by TEV, indicating that the protein might not have been completely aggregated, another 12 l large scale expression and purification of this protein was performed. The first two elution fractions from the GST-affinity column were pooled and digested with TEV-protease (Figure 33 a, b). The success of the digestion was verified by SDS-PAGE and the digested protein was separated on an S200 16/60 column (Figure 33 d). VirGK3 was identified in the void again, the only soluble protein was the removed GST-tag and the TEV-protease, corresponding to the lower and the upper band in the green fractions respectively (Figure 33 c).

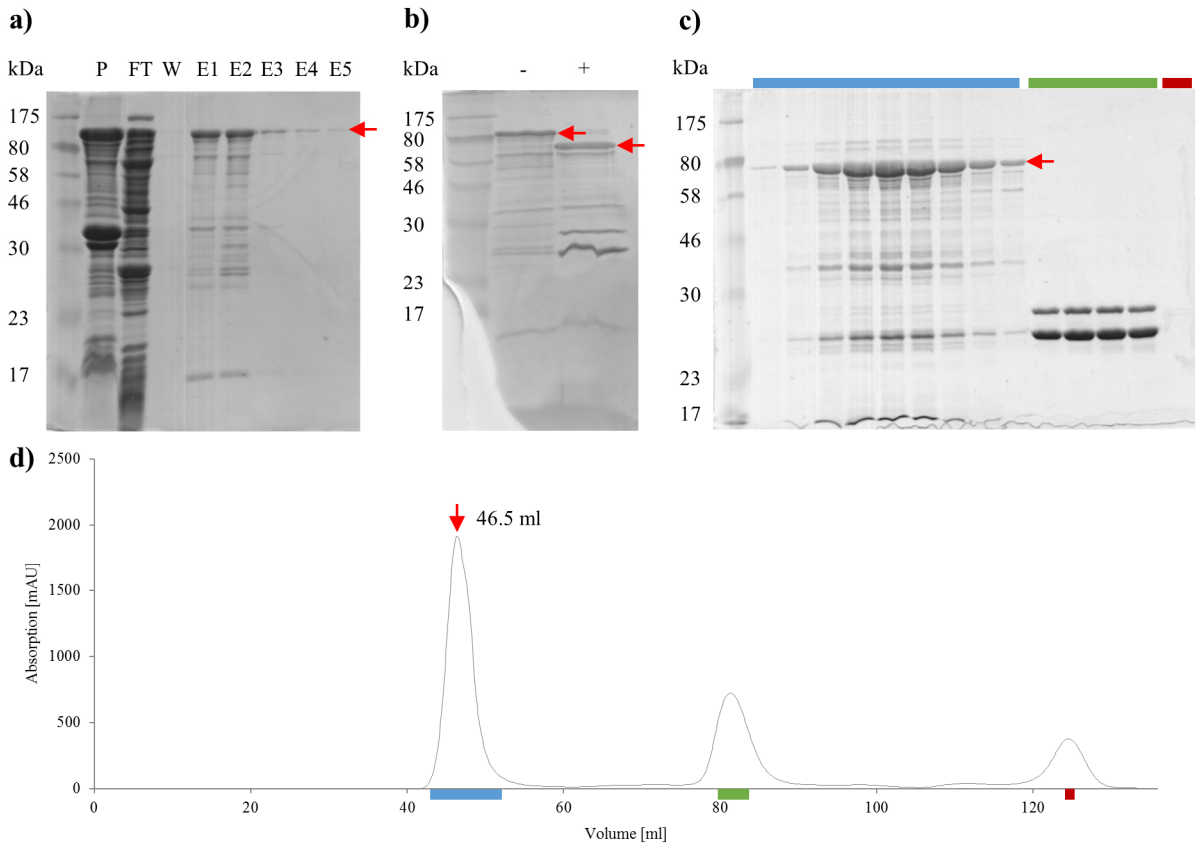


Figure 33: GST-VirGK3 large scale expression and purification. As a first step, the protein was purified, using a GST-affinity column. In a second step the protein mixture was separated, using an S200 16/60 column. a) Coomassie stained SDS-gel of different steps in the GST-affinity purification: P) pellet (insoluble protein), FT) flow through of the column, W) wash, E1-E5) elution fractions (20 mM GSH); b) Coomassie stained SDS-gel to control TEV-protease cleavage: -) control without protease, +) cleaved protein; c) Coomassie stained SDS-gel of different elution fractions from the S200 16/60 column. Elution fractions are color coded and correspond to the fractions marked in d); d) Elution profile of the run on the S200 16/60 column. Elution volumes are indicated. Red arrows mark VirGK3.

To further investigate if VirGK3 was really aggregating or just formed a very large, but uniform complex, running in the void of an S200 column, a part of the eluted protein was rerun on a Superose6 10/30 column that is able to separate proteins and complexes of up to 2 MDa (Figure 34). VirGK3 also eluted in the void volume of the Superose6 column, hence it formed soluble aggregates of sizes larger than 2 MDa.

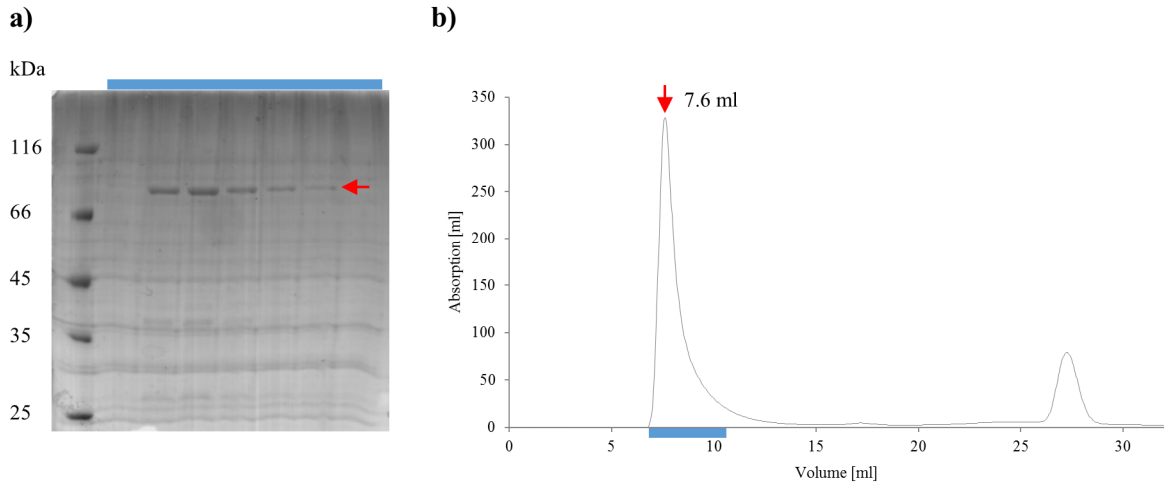
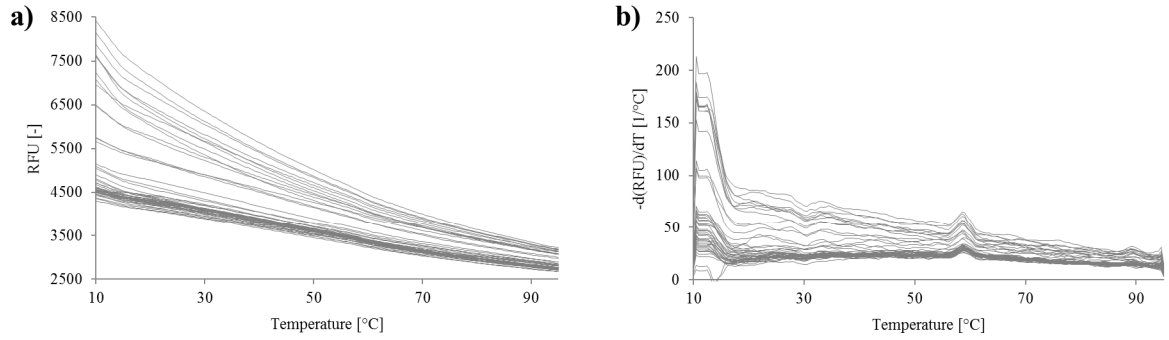


Figure 34: VirGK3 run on a Superose6 10/30 column. a) Coomassie stained SDS-gel of different elution fractions from the Superose6 column. Elution fractions are color coded and correspond to the fractions marked in b); b) Elution profile of the run on the Superose6 column. Elution volumes are indicated. Red arrows mark VirGK3.

As VirGK3 formed soluble aggregates, a thermofluor-assay was performed, using the purified protein, to test for better buffer conditions. However, no distinct melting curves were observed (Figure 35), indicating that the protein wasn't properly folded in the first place. The SYPRO Orange fluorophore was able to bind to exposed hydrophobic patches right from the beginning, resulting in high initial fluorescence, which decreased as the protein aggregated with rising temperature. Hence no better buffer conditions and no stabilizing additives were identified for VirGK3.

Buffer screen:



Additive screen:

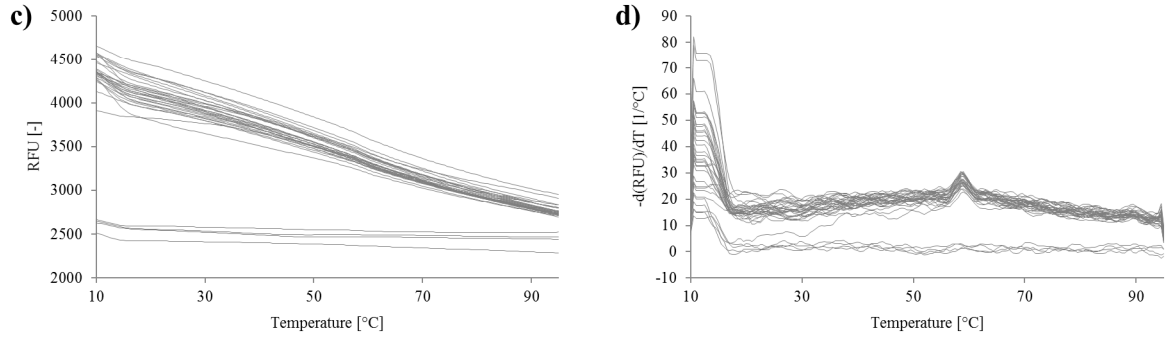


Figure 35: Thermofluor-assay for VirGK3. a) Melting curves and b) melting peaks of a buffer screen. c) Melting curves and d) melting peaks of an additive screen. No distinct melting curves were observed, indicating that the protein wasn't properly folded in the first place.

VirGK3 formed soluble aggregates, which was shown by SEC and in a thermofluor-assay. To test for the solution state properties of the other VirG-constructs, in the next step GST-VirGK1-K4 were expressed and purified in large scale. The protein containing elution fractions from the GST-affinity column were pooled and split in half. One half was digested with TEV to remove the GST-tag, the other half was not, to test if the solubility tag could prevent protein aggregation. All constructs were separated on a Superose6 column and all constructs, with or without GST, eluted in the void volume, indicating that all of them formed soluble aggregates.

Up to this point several VirG-constructs were expressed and purified, but all of them formed soluble aggregates. It was likely that the IcsB/ATG5 binding motif was not accessible in this form and that the protein was not able to bind its interaction partners. Additionally it was improbable that the protein crystallized in this state and that its structure could be solved. Hence at this point it was necessary to try a different approach for the identification of better constructs or conditions for expression and purification, to obtain suitable protein for crystallization and interaction studies.

4.1.2 The ESPRIT system

As described in chapter 3.2.7 the ESPRIT system is a system to screen for soluble constructs of difficult to express proteins, which was applied to VirG at this point. Two different C-termini were selected, residues 592 and 758. 592 and 758 were the approximate boundaries of the autochaperone domain, as identified by limited proteolysis [133]. All constructs ending with residue 592 excluded the autochaperone domain, while the constructs ending with residue 758 included it. This approach was chosen to make sure that not only different constructs of the autochaperone domain were identified, which had already been shown to be expressed solubly.

In preparation of the experiment, for each of the two C-termini two libraries of N-terminally truncated constructs were created, one containing smaller and one containing larger fragments, resulting in a total of four libraries named 592(S)mall/(L)arge and 758 S/L. This preparatory step was performed by Philippe Mas in the laboratory of Dr. Darren Hart at the EMBL in Grenoble, France.

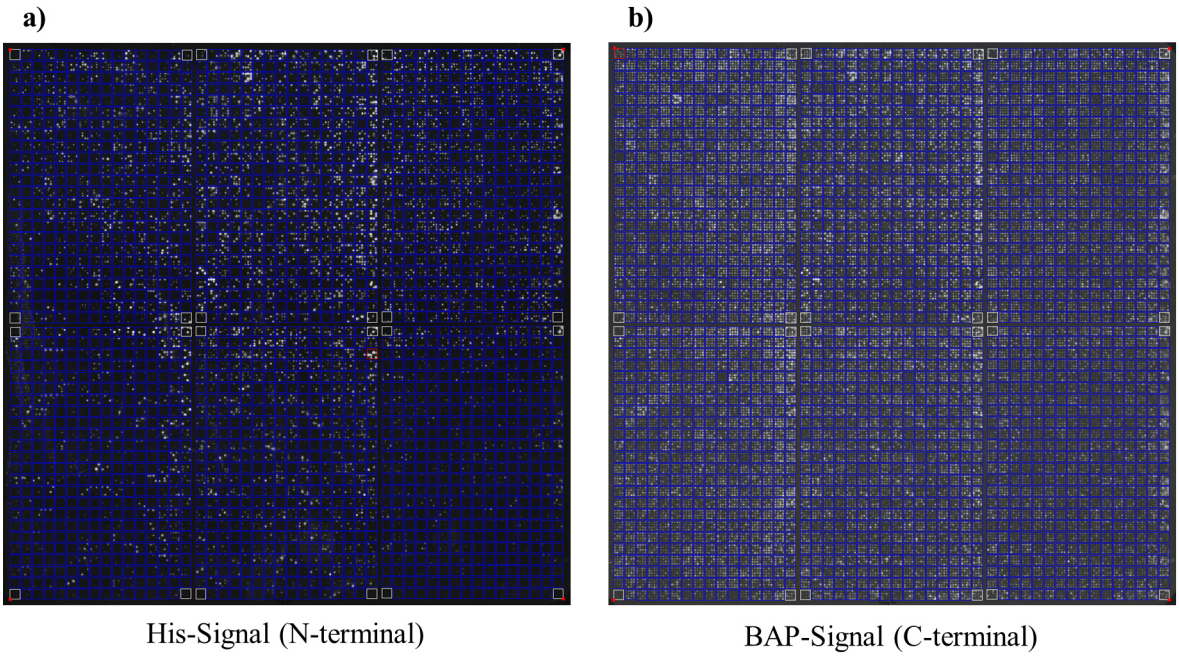


Figure 36: Fluorescence scan of the nitrocellulose membrane, hybridized with a fluorescently labeled anti-His antibody and a fluorescently labeled streptavidin to detect the presence of a) the N-terminal His-tag and b) the C-terminal BAP. The pattern, by which the colonies were blotted onto the membrane, is described in chapter 3.2.7.

The following steps were performed during a three week visit in the laboratory at the EMBL Grenoble, where Philippe Mas and I performed the experiments together. From the four libraries a total of 18400 clones were picked (592S: 3800, 592L: 4600, 758S: 3500, 758L: 6500 clones) and screened for soluble protein expression. They were gridded on a nitrocellulose membrane and the protein was detected by a Western blot like method, using fluorescently labeled Anti-His antibody and streptavidin, to label the N-terminal His-tag and the C-terminal BAP with fluorescent probes. The membrane was scanned, using a fluorescence scanner afterwards (Figure 36).

Based on the fluorescence profile, the 96 best clones were selected, 24 clones per library. One clone was replaced by a positive control, expressing MBP, for the following expression tests. The expression was tested in 4 ml of TB. Samples of the purified protein were run on an SDS-gel and analyzed by Western blotting, using an anti-His primary antibody (Figure 37). The Western blot showed that many clones in the two 758 libraries expressed a protein with a MW of around 20 kDa. Most probably this fragment corresponded to the autochaperone domain, which has a MW of 18 kDa. This was a first hint that soluble domains could be identified. Clones showing higher MW bands were of special interest. Positive results were obtained for VirGE57, E76, E81, E82 and E94. Also in the two 592 libraries many clones showed a band with a size of around 20 kDa. As the autochaperone domain was not present in these constructs, the sequencing had to show which the boundaries of these constructs were. Also in these libraries some proteins with higher MW bands were identified, especially VirGE5, E14-E16, E20 and E21.

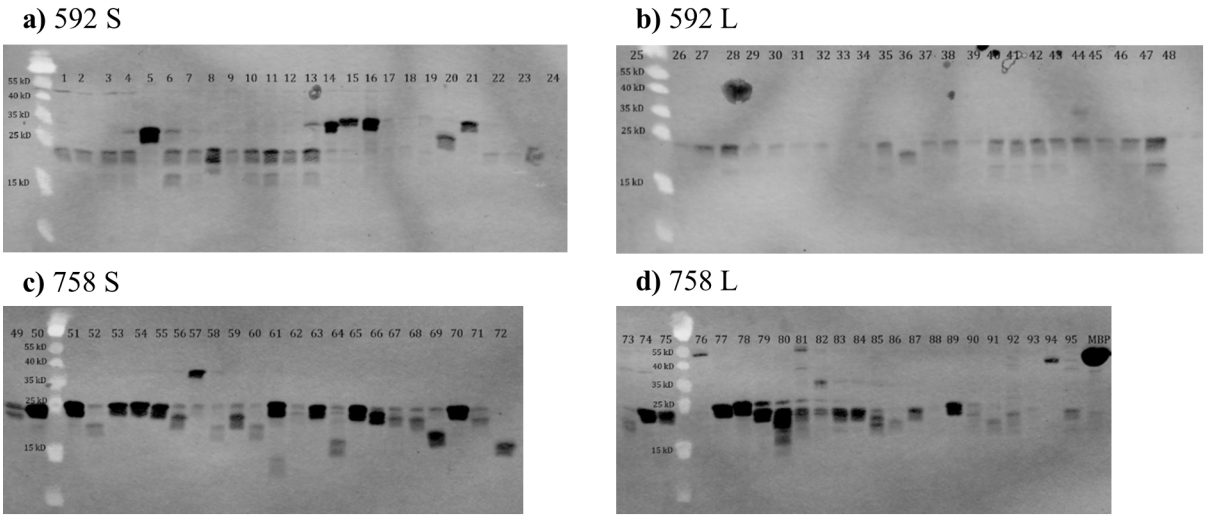


Figure 37: Western blots of the best 24 clones of each library, 95 clones and one positive control (MBP) in total, using an anti-His primary antibody. Many clones in the two 758 libraries showed a band at a size of around 20 kDa, which corresponds to the size of the autochaperone domain (18 kDa). Some also had a higher MW e.g. VirGE57, E76, E81, E82 and E94. Also for the two 592 libraries many clones showed a band at around 20 kDa, but some also had a higher MW e.g. VirGE5, E14-E16, E20 and E21. Apparent double bands are SDS-PAGE migration artifacts.

Out of the 96 clones, 27 were selected that showed strong bands in the Western blot or interesting MWs. Samples of the purified proteins were analyzed on a Coomassie stained SDS-gel to assess the protein quantity (Figure 38). Most of the clones, expressing a protein of a size of around 20 kDa, showed strong expression, while higher MW constructs were expressed more weakly. As the scale up of weakly expressing constructs, identified by this method, is sometimes challenging, it was uncertain if the expression and purification of these constructs would have been possible.

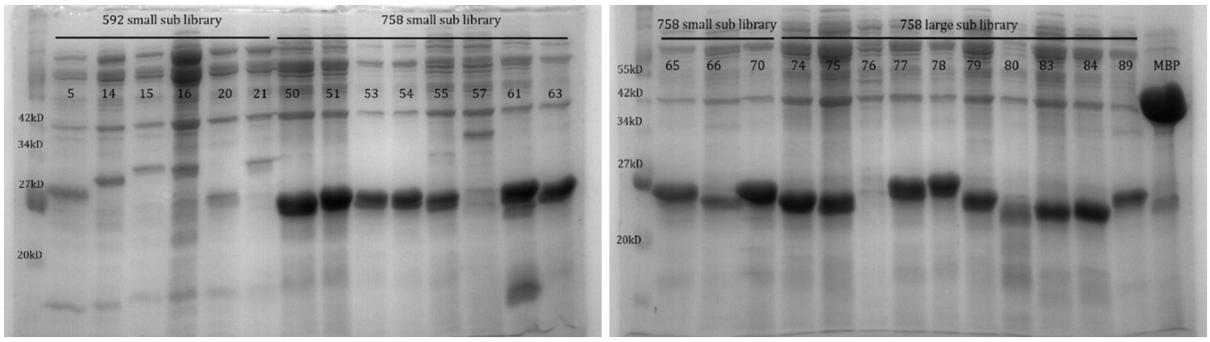


Figure 38: Coomassie stained SDS-gels of 27 selected clones and the positive control (MBP). Most clones, showing a band at around 20 kDa, showed good expression, while the clones with higher MW bands showed weaker expression.

The next step was the sequencing of the 95 clones from the N-terminal side, as the C-termini were fixed. Only the clones, showing interesting N-terminal boundaries, were sequenced completely. In Table 15 the results of the sequencing reactions are shown. A number of clones contained contaminating DNA sequences (cont.), i.e. DNA sequences differing from *virG*. Other clones contained constructs of the autochaperone domain, which was a sign that the method worked, as the autochaperone domain previously had been shown to be a good expressing VirG-construct. Finally 18 interesting constructs have been identified that covered parts of the VirG protein, where no structural information was available yet. 4 of the 18 constructs contained a frame shift mutation in the beginning, making the protein expression questionable, even if it was shown in small scale test-expressions. 13 of the 18 constructs covered the IcsB/ATG5 binding domain at least partially (Figure 39).

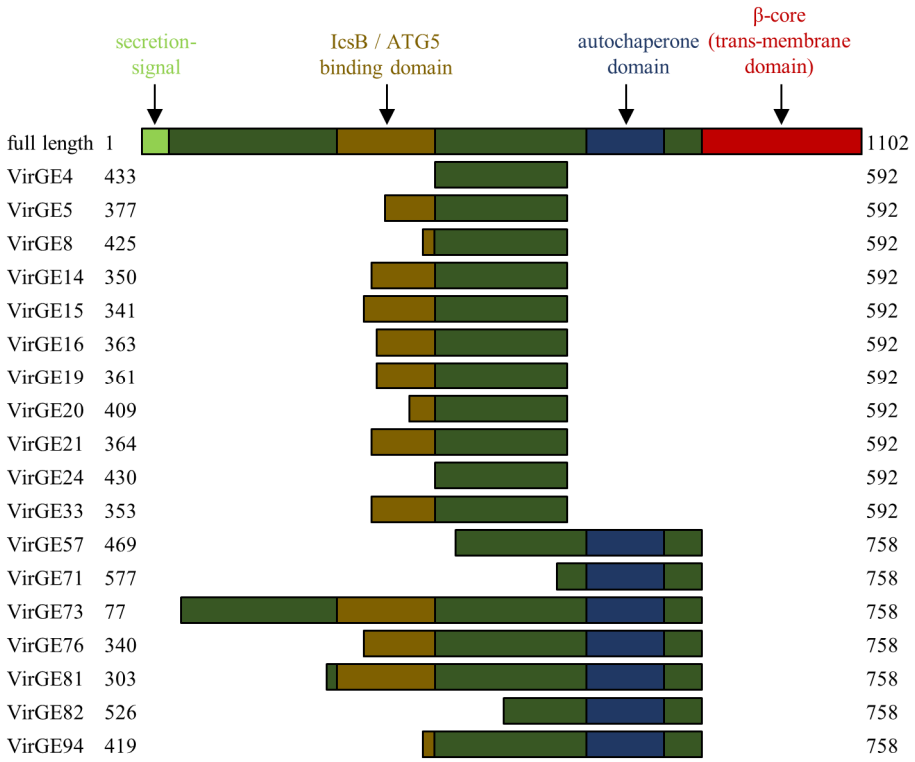


Figure 39: Overview of the 18 interesting VirG-constructs from the ESPRIT screening.

Table 15: Sequencing results of the 96 best clones from the ESPRIT system. The coding sequence column indicates which coding sequence has been identified in the vector (compare text). Start and end indicate the boundaries of the respective construct, MW the molecular weight of the resulting protein. A violet background indicates a construct similar to the already known autochaperone domain, a green background indicates an interesting construct that covered unknown regions of VirG and a yellow background indicates an interesting construct, containing a frame-shift mutation in the beginning.

592 S

	coding sequence	start	end	MW [kDa]
1	cont.			
2	cont.			
3	cont.			
4	VirG	433	592	17,1
5	VirG	377	592	22,8
6	cont.			
7	cont.			
8	VirG	425	592	18
9	cont.			
10	cont.			
11	cont.			
12	cont.			
13	cont.			
14	VirG	350	592	25,7
15	VirG	341	592	26,5
16	VirG	363	592	24,2
17	cont.			
18	cont.			
19	VirG	361	592	24,5
20	VirG	409	592	19,7
21	VirG	364	592	24,1
22	cont.			
23	cont.			
24	VirG	430	592	17,5

592 L

	coding sequence	start	end	MW [kDa]
25	cont.			
26	cont.			
27	cont.			
28	cont.			
29	cont.			
30	cont.			
31	cont.			
32	cont.			
33	VirG	353	592	25,3
34	cont.			
35	cont.			
36	VirG	629	758	
37	cont.			
38	cont.			
39	cont.			
40	cont.			
41	cont.			
42	cont.			
43	cont.			
44	cont.			
45	cont.			
46	cont.			
47	cont.			
48	cont.			

758 S

	coding sequence	start	end	MW [kDa]
49	VirG	614		
50	VirG	604	758	16,6
51	VirG	597		17,3
52	VirG	634		
53	VirG	597		
54	VirG	597		
55	VirG	601		
56	VirG	626		
57	VirG	469	758	31,1
58	VirG	638		
59	VirG	618		
60	VirG	638		
61	VirG	614		
62	cont.			
63	VirG	601		
64	VirG	592		
65	VirG	604		
66	VirG	609		
67	VirG	618		
68	VirG	626		
69	VirG	659		
70	VirG	604		
71	VirG	577	758	19,5
72	VirG	667		

758 L

	coding sequence	start	end	MW [kDa]
73	VirG	77	758	69,7
74	VirG	612		15,7
75	VirG	617		15,2
76	VirG	340	758	44,4
77	VirG	601		
78	VirG	597		17,3
79	VirG	612		15,7
80	cont.			
81	VirG	303	758	48,2
82	VirG	526	758	24,7
83	cont.			
84	VirG	617		
85	VirG	629		
86	VirG	636		
87	VirG	619		
88	VirG	638		
89	VirG	601		
90	VirG	619		
91	VirG	<653	758	
92	VirG	619		
93	cont.			
94	VirG	419	758	36,5
95	VirG	617		
96	MBP			

As a next step, test-expressions in a one liter scale were done for most of the 18 interesting ESPRIT-constructs, as with regard to the small scale test-expressions only weak protein expression levels were expected. Test-expressions were performed and the elution fractions from the His-affinity column were analyzed on an SDS-gel (Figure 40). More than half of the constructs showed, in most cases faint, protein expression. The better expressing constructs were VirGE57 and VirGE94 and even there the bands were weak. As a first test of the properties of these constructs, VirGE57 was separated on a Superose6 column, where it eluted in the soluble fraction, indicating that it didn't form aggregates.

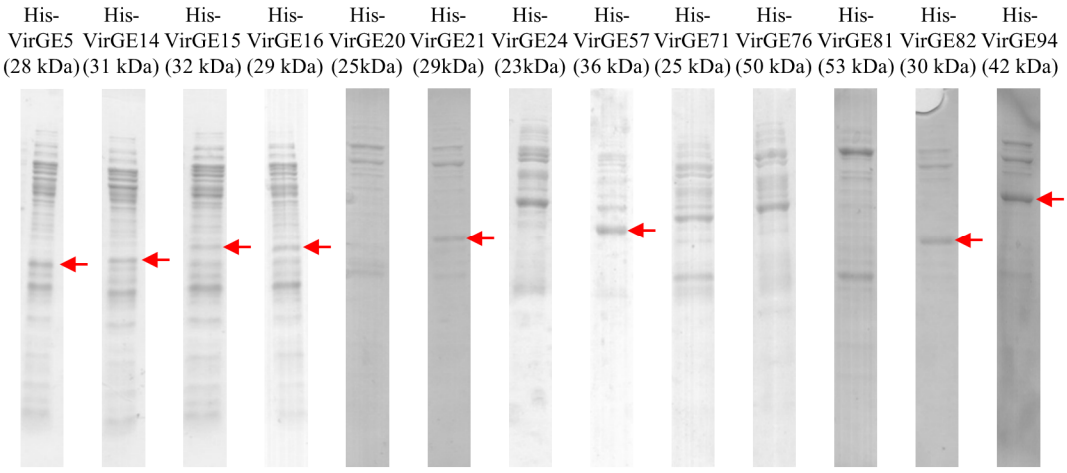


Figure 40: Coomassie stained SDS-gels of large scale test-expressions of His-VirG-constructs. Only soluble protein expression is shown. Red arrows mark the corresponding VirG protein band if present. The best expression was observed for VirGE57 and VirGE94.

Consequently large scale expressions and purifications on a scale of 8 to 16 liters were done for VirGE57 and two additional constructs. VirGE94 and VirGE76 were chosen, because they comprised some of the longest constructs, which resulted from the ESPRIT screening. VirGE94 showed a band in the test-expression, while VirGE76 didn't, but it contained most of the IcsB/ATG5 binding domain. Large scale expressions of all three constructs resulted in solubly expressed protein (Figure 41), even for the VirGE76. The highest yields were obtained for VirGE94 (about 1.1 mg of purified protein per liter of TB), while VirGE76 yielded the least protein (about 0.2 mg of purified protein per liter of TB).

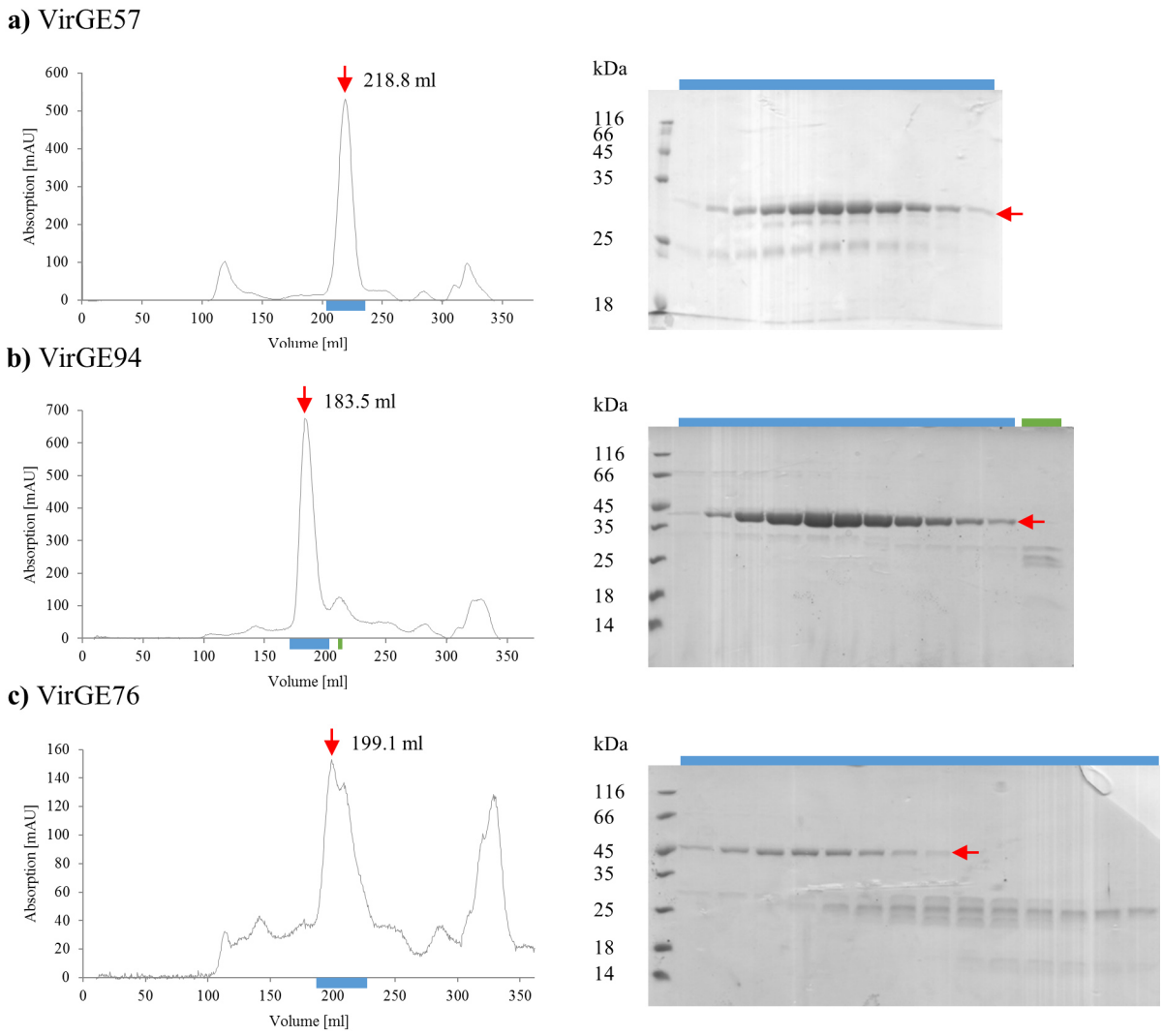


Figure 41: Large-scale expressions and purifications of a) VirGE57, b) VirGE94 and c) VirGE76. As a first step the protein was purified, using a His-affinity column. In a second step the protein mixture was separated, using an S200 26/60 column. Elution profiles and Coomassie stained SDS-gels of the runs on the S200 26/60 column are shown. Elution fractions are color coded in the elution profiles and the gels. Elution volumes are indicated. Red arrows mark the corresponding VirG-proteins.

Also the expression of other VirG-constructs has been tested in large scale expressions. Interestingly it was not possible to obtain soluble protein for constructs with the C-terminal residue 592, lacking the autochaperone domain. As an example Figure 42 shows an SDS-gel with samples from the His-affinity purification of VirGE14. All protein expression was insoluble in the pellet, no soluble protein was eluted from the His-affinity column.

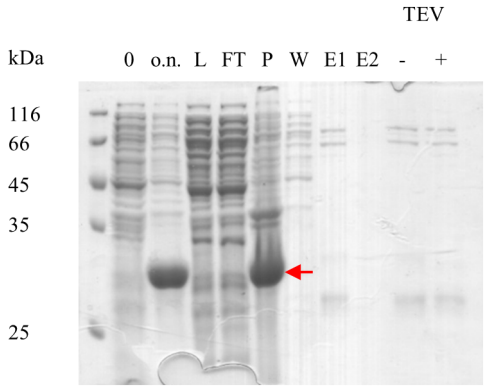


Figure 42: VirGE14 large-scale expression and purification. Coomassie stained SDS-gel of different steps in the His-affinity purification: 0, o.n.) samples taken directly from the culture before induction and after overnight expression, L) protein that was loaded to the column, FT) flow through of the column, P) pellet (insoluble protein), W) wash, E1) elution fraction (250 mM imidazole), E2) elution fraction (500 mM imidazole). -TEV) cleavage control without protease. +TEV) cleaved protein. The red arrow marks VirGE14. No soluble protein expression was observed.

It was tried to improve the soluble protein yields of the insolubly expressed VirG-constructs, as many of them showed protein expression in the small and large scale test-expressions. VirGE14 was used to test different culture and lysis conditions. It was expressed and lysed, using different lysis buffers, containing additives that had been shown to have a stabilizing effect (10% Mannitol, 10% L-Arginine, 10% NDSB 201, 10% K_2HPO_4, 10% CTAB, 10% $CuCl_2$) [187]. Also addition of different concentrations of detergent was tested (0.03-3% DDM), as well as cell lysis using 10% Bugbuster, which also contained detergents. Only upon addition of very high concentrations of DDM (3%), a faint band of soluble VirGE14 was observed. Expression at very low temperatures (4°C) had been shown to improve the expression of some proteins [188], but after three to four days of expression at 4°C only an increase in the insolubly expressed protein was observed. Expression of VirGE76 at 4°C resulted in the disappearance of soluble protein expression. In summary, these methods failed to improve the soluble protein yields of the insolubly expressed VirG-constructs.

The solubly expressed protein VirGE57 and VirGE94 were used for a thermofluor-assay to test for better buffer conditions for expression and purification (Figure 43). VirGE57 showed typical melting curves but the VirGE94 melting curves were unusual, showing very high and constant fluorescence values in the beginning and a sudden drop in fluorescence shortly before the melting temperature. The reason for this behavior was not known. Both proteins had melting temperatures of around 55°C. The ideal buffer conditions for both constructs were similar (50 mM HEPES pH 7.0, 300 mM NaCl) and close to the buffer that had already been used. Consequently the buffer wasn't changed.

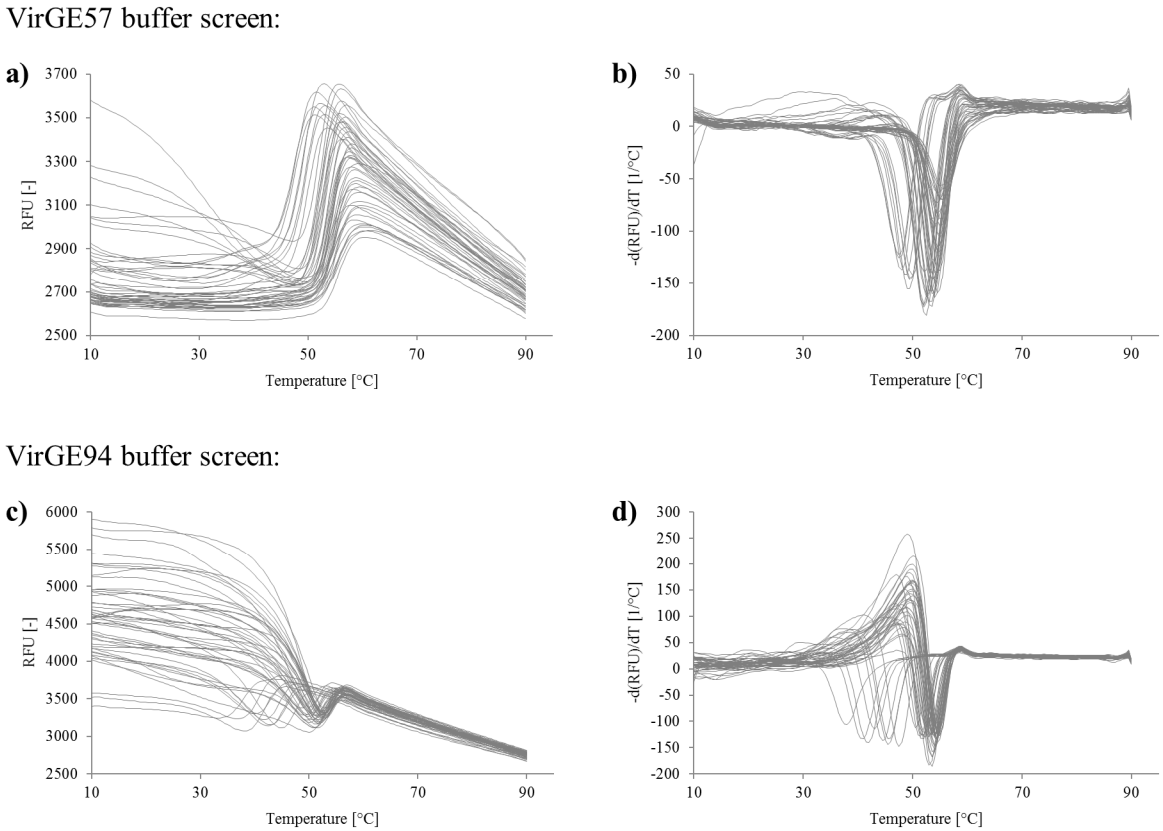


Figure 43: Thermofluor-assay for VirGE57 and VirGE94. a) Melting curves and b) melting peaks of a buffer screen for VirGE57. c) Melting curves and d) melting peaks of a buffer screen for VirGE94. Both proteins had melting temperatures of about 55°C, but the melting curves for VirGE94 were unusual.

4.1.3 VirGE94 crystallization and structure

Crystallization setups were done for all solubly expressed VirG-constructs (VirGE57, E94, and E76). VirGE57 and VirGE94 showed initial hits in different crystallization screens. The most promising was obtained for VirGE94 in condition 66 of the Midas screen (0.2 M KCl, 0.1 M TRIS pH 8.0, 30% Jeffamin M-2070) at 20°C, where small three dimensional crystals appeared after two days and continued growing for at least three weeks (Figure 44 a). The crystallization condition was reproduced, using a custom 24-well grid screen around this condition. Initially a shower of small crystals was obtained (Figure 44 b). When the crystals grew larger, additional smaller crystals began to grow on their surface, leading to multiple crystals growing into each other, resembling an intertwining spiral with a solid core (Figure 44 c). Several of those crystals were dissolved in SDS loading buffer and analyzed on an SDS-gel to identify the protein (Figure 44 e). The size corresponded to VirGE94 (37 kDa) and the protein in the band was identified as VirG by tryptic digestion and mass spectrometry. Finally a large crystal with dimensions of around 130x100x50 µm was found that was composed only of the solid core without crystals growing on top of it (similar to Figure 44 d). This crystal was used for data collection.

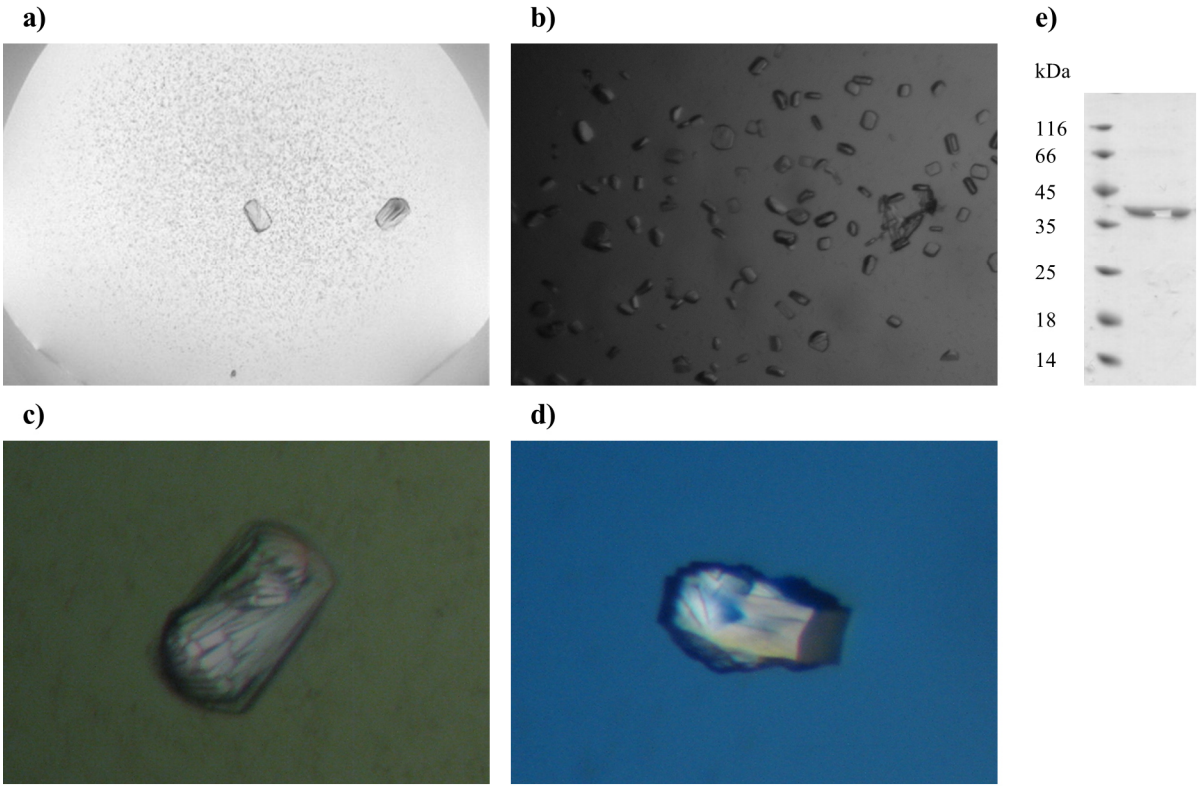


Figure 44: VirGE94 crystals. a) Condition 66 from Midas screen (0.2 M KCl, 0.1 M TRIS pH 8.0, 30% Jeffamin M-2070, 5 mg/ml, 20°C). b) Reproduced in 24-well hanging drop setups (0.2 M KCl, 0.1 M TRIS pH 7.0, 26.7% Jeffamin M-2070, 5 mg/ml, 20°C). c) Under optimized conditions (0.2 M KCl, 0.1 M TRIS pH 7.8, 26.7% Jeffamin M-2070, 5 mg/ml, 20°C) the crystals grew to a larger size, but smaller crystals grew on top of the larger ones. d) Finally some large crystals with few or no crystals on top and dimensions of up to 130x100x50 µm were found (0.2 M KCl, 0.1 M TRIS pH 7.2, 26% Jeffamin M-2070, 2.5 mg/ml, 20°C). e) The size of the protein in the crystal corresponded to the size of VirGE94 and the protein was identified as VirG by tryptic digestion and mass spectrometry.

The crystal diffracted to a resolution of around 3.0 Å in house. A dataset was collected at beamline 14.1 of the BESSY II synchrotron, using the installed Pilatus 6M detector, an X-ray wavelength of 0.918 Å and a temperature of 100 K. The crystal diffracted to a resolution of around 1.9 Å there. The data was integrated using XDS and scaled using XSCALE. The structure was solved by molecular replacement using Phaser-MR and the published structure of the VirG autochaperone domain (PDB-entry 3ML3) as a search model. Phenix AutoBuild was used for initial automated model building. WinCoot and Phenix.refine were used for later manual model building and refinement steps.

VirGE94 crystallized in the $P2_1$ space group with 2 molecules in the asymmetric unit. The structure of VirGE94 is shown in Figure 45. The two molecules in the asymmetric unit were very similar to each other with a RMSD of 0.25 Å for all atoms and 0.23 Å for the backbone residues. Minor differences were seen primarily in the conformation of the first N-terminal residues.

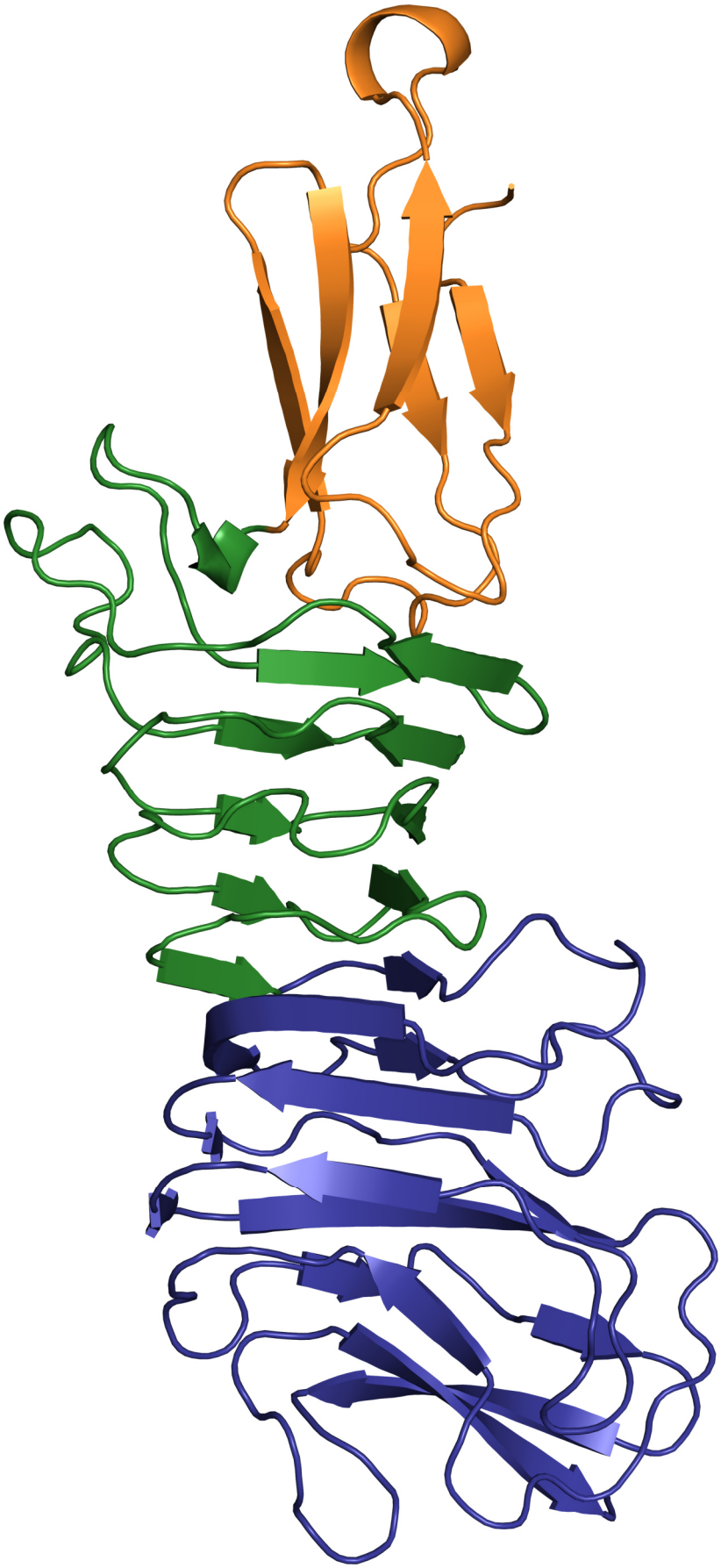


Figure 45: Crystal structure of VirGE94. Blue: autochaperone domain 591-740 (already known), green: continuation of the β-helix 484-590, orange: new domain of unknown function (VirG insertion domain/VID) 422-483.

Table 16: Data collection and refinement statistics for VirGE94.

Data collection	
X-ray source	BM 14.1, BESSY II
Wavelength [Å]	0.918
Resolution [Å]	1.90 (1.97 – 1.90)
Space group	$P2_1$
Unit-cell parameters [Å]	35.9, 84.7, 114.3 90.0, 98.5, 90.0
R_{merge}	0.091 (0.817)
R_{meas}	0.099 (0.884)
CC1/2	0.999 (0.824)
‹I/σ(I)›	13.11 (2.50)
Completeness [%]	98.5 (97.6)
No. of reflections	360053 (35865)
Multiplicity	6.8 (6.9)
Refinement	
Resolution [Å]	42.3 – 1.90
No. of reflections	52462 (5172)
R_{work}	0.167 (0.252)
R_{free}	0.211 (0.285)
Asymmetric unit	2 molecules
No. of atoms	
Protein	4829
Ligands	5
Water	420
Average *B* factor [Å^2]	
Protein	40.4
Ligands	58.9
Water	41.9
R.m.s. deviations	
Bond length [Å]	0.006
Bond angles [°]	0.97
Ramachandran favored [%]	97
Ramachandran outliers [%]	0.3

The structure of VirGE94 was composed of the C-terminal autochaperone domain, comprising residues 591-740 (blue in Figure 45), an intermediate part, comprising residues 484-590, that extended the β-helix by about four turns (green in Figure 45) and an N-terminal region, comprising residues 422-483, that formed a previously unknown domain (orange in Figure 45), which is named VirG insertion domain (VID) hereafter.

One interesting property of VirGE94 in comparison to VirGE57 and VirGE76 was, that it formed dimers in solution. As can be seen in Figure 41, it eluted earlier from a SEC column than the larger VirGE76. In the crystal structure two different contacts that could have been responsible for the dimerization were identified. First the VID was involved in a crystal contact by the formation of an intermolecular β-sheet with a symmetry related copy of itself, which was rotated by 180°. The second contact was mediated by a Ni^{2+}-ion, which was stripped off the His-affinity column during purification.

An experiment was designed to test if the contact, mediated by the Ni^{2+}-ion was responsible for the dimerization of VirGE94. VirGE94 was expressed and purified on a His-affinity column. The eluted protein was split in half. One half was dialyzed overnight against normal buffer, the other half against buffer containing 1 mM EDTA. The next day both samples were separated on an S200 10/30 column (Figure 46). If the Ni^{2+}-ion was responsible for the dimerization, the VirGE94 dimer should have been disrupted in the preparation with EDTA, as EDTA depletes the Ni^{2+}-ions by complexing them. The different size of the resulting monomers should have been detectable by SEC. However, no significant differences in the elution profile or volume were detected. Samples of the protein containing fractions

were analyzed on an SDS-gel and also looked similar. Hence the VirGE94 dimer was not disrupted by the addition of EDTA. There were two possible explanations for this. Either the dimer, present in solution, was not the one mediated by the Ni^{2+}-ion, or the EDTA was not capable of removing the ion from the dimer interface. As the Ni^{2+}-ion seemed to be quite accessible in the crystal structure (Figure 113) and the affinity of Ni^{2+} to EDTA was most probably higher than to VirGE94, the first explanation was more probable.

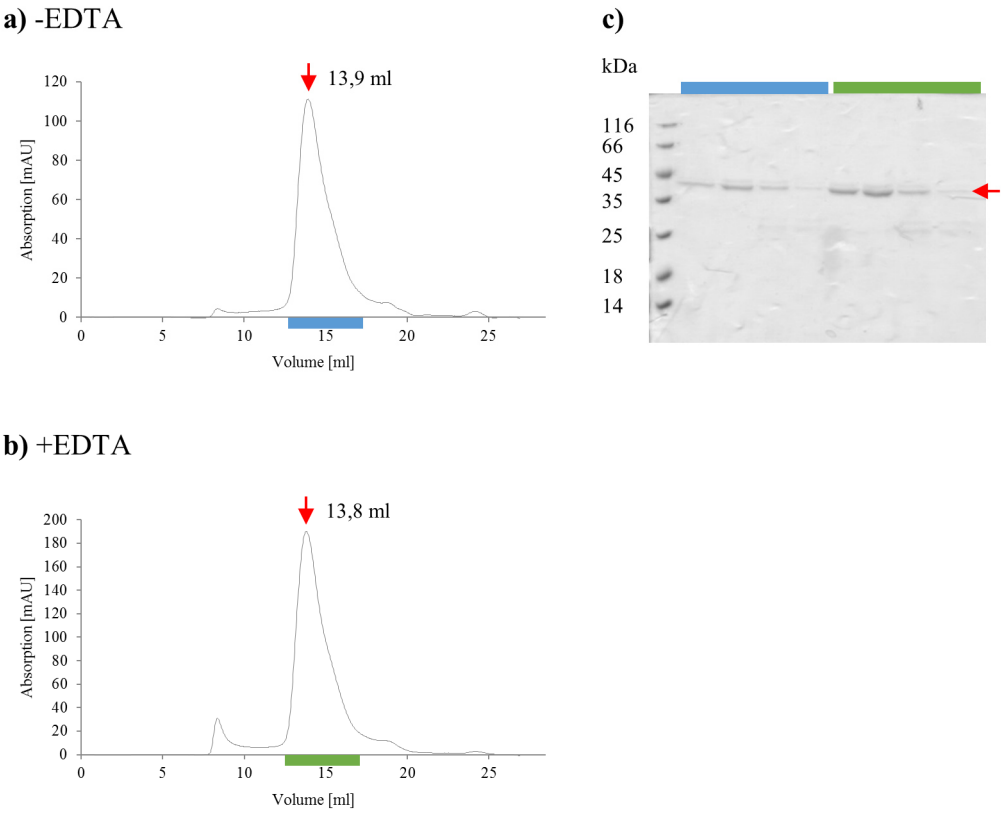


Figure 46: VirGE94 dialyzed against buffer a) without or b) with EDTA, separated on an S200 10/30 column. Elution profiles and c) a Coomassie stained SDS-gel of the two runs are shown. Elution fractions are color coded in the elution profiles and the gel. Elution volumes are indicated. Red arrows mark VirGE94. No significant differences were observed in the two preparations of the protein.

To test if the other crystal contact, mediated by the VID, was responsible for the dimerization of VirGE94 in solution, SAXS experiments were performed. By comparing the theoretical scattering curves of the different possible crystallographic dimers with the experimental scattering data, it should have been possible to determine, which one was the dimer, present in solution. Another question that was addressed by this experiment was, where the VID was located in the full length protein. It might have occupied the same position, as in the crystal structure or it might have been attached to the side of a continued β-helical stem. It should have been possible to locate this domain, by calculating SAXS envelopes from the experimental scattering data.

SAXS experiments were performed, using VirGE57, VirGE76, VirGE94 and two additional constructs: VirGK26 and VirGK34 (Figure 47 a). The origin of VirGK26 and VirGK34 is described in chapter 4.1.4. The expected solution states of the constructs are shown in Figure 47 b. Of special interest was the VirGE76 construct, as it should have been possible to determine the position of the VID in the full length protein from its envelope.

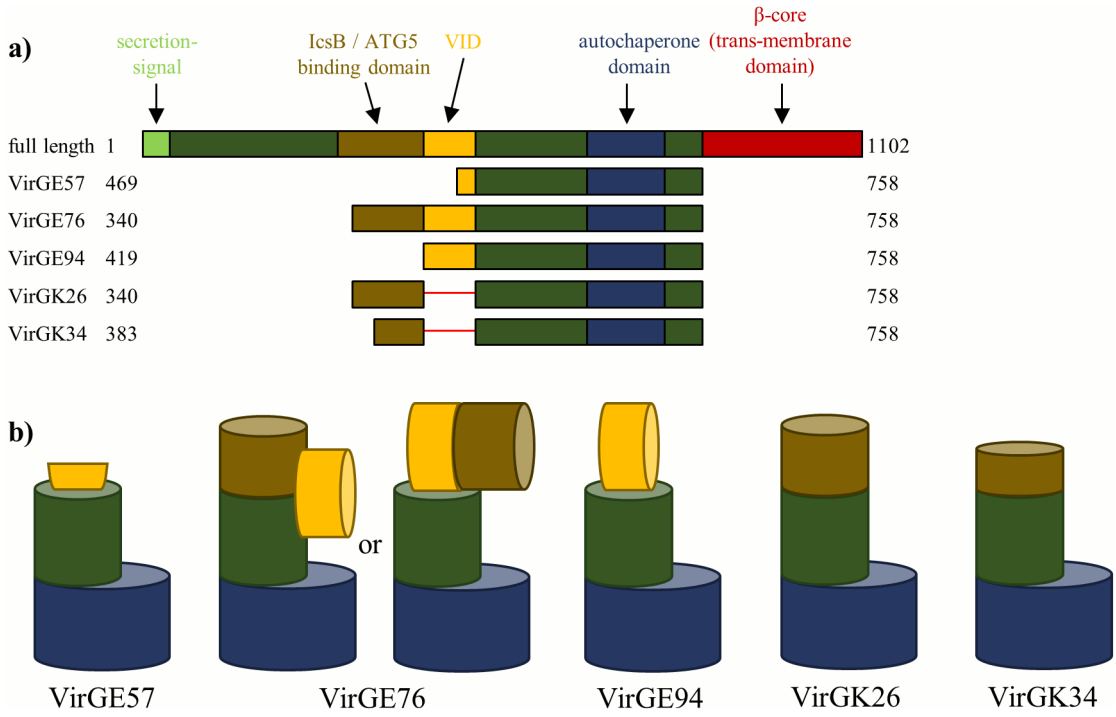


Figure 47: a) Overview of the VirG-constructs used for SAXS experiments. b) Expected solution states of the constructs.

The scattering curves and distance distribution functions of all five proteins are shown in Figure 48. The distance distribution functions of all constructs showed a bi- to multi modal distribution, which indicated that these proteins formed aggregates. The SAXS data were used to calculate the MW of the VirG-constructs, which are shown in Table 17. VirGE57 was clearly monomeric and VirGE94 formed a dimer, which confirmed the results from SEC. VirGK34 also seemed to be dimeric in solution, while VirGE76 and VirGK26 formed higher oligomers, which were most probably soluble aggregates. VirGE76 was clearly monomeric, when it eluted from the SEC column and VirGK26 was either a monomer or a dimer. Most probably the proteins aggregated, while being transported to the synchrotron for data collection, as the distance distribution functions of all proteins showed signs of aggregation (Figure 48).

Table 17: MWs and the corresponding oligomerization states of the different VirG-constructs, calculated from the SAXS data.

Construct	Molecular weight [kDa]	Oligomerization state
VirGE57	38.2	Monomer
VirGE94	88.2	Dimer
VirGE76	157.1	Aggregate?
VirGK26	541.9	Aggregate
VirGK34	51.3	Dimer

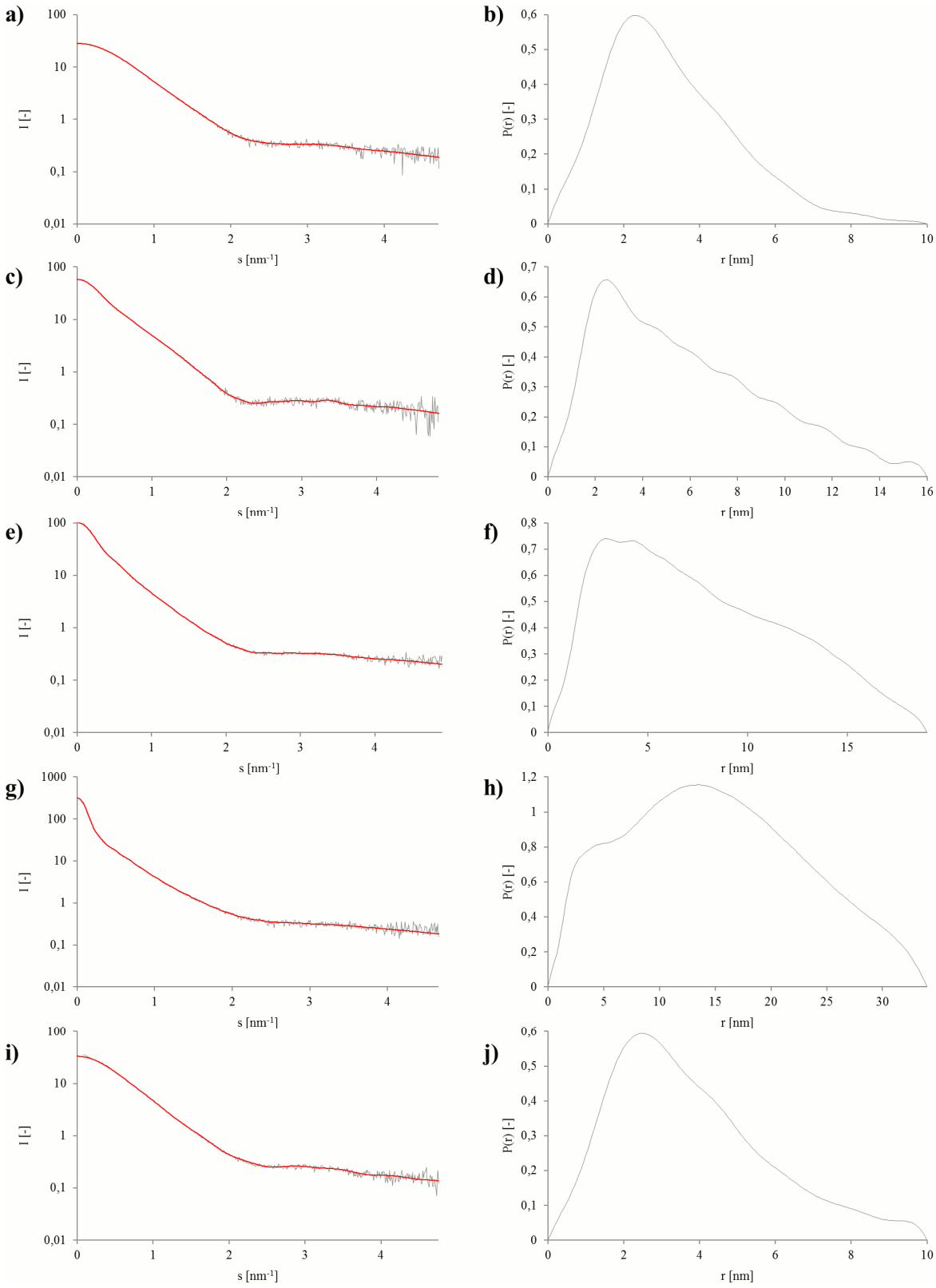


Figure 48: (a), (b) VirGE57, (c), (d) VirGE94, (e), (f) VirGE76, (g), (h) VirGK26 and (i), (j) VirGK34 have been used in SAXS experiments. In the left column experimental (grey) and regularized (red) intensities are plotted against the scattering length s. In the right column particle distance distribution functions P(r) are shown.

To identify the dimer interface that was responsible for the solution state of VirGE94, the theoretical SAXS scattering curves of all possible crystallographic dimers were calculated and fitted to the experimental data. The dimer that was formed by the interaction of the VIDs resulted in the best fit with a χ^2-value of 6.7 (Figure 49). The fit was good at lower scattering length, corresponding to lower resolutions, but deviated distinctly at higher scattering length and resolutions, indicating that the solution structure differed in details from the crystal structure.

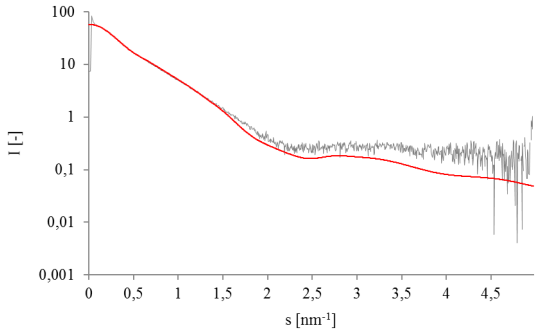


Figure 49: Fitting of the calculated SAXS scattering curve of the VirGE94 dimer, formed by the interaction of the VIDs (red), to the experimental scattering curve (grey). For the fit a χ^2-value of 6.7 was obtained.

Envelopes were calculated for VirGE57, VirGK34 and VirGE94 and models of the proteins were fitted into these envelopes (Figure 50). The VirGE94 envelope showed an elongated form, confirming the dimer formation *via* the VIDs, as also this dimer formed a long rod (Figure 50 c). A truncated model of VirGE94 was used to fit the experimental data of VirGE57 (Figure 50 a). No model was present for VirGK34, as it was not known if the additional N-terminal amino acids adopted a β-helical conformation. Thus a truncated VirGE94 model was used, lacking the additional N-terminal amino acids (Figure 50 b).

All envelopes possessed a “tail”, which at least for VirGE94 was most probably an artifact of the calculation of the envelope, indicating that the protein was aggregated (personal communication with Dmitri Svergun from the EMBL Hamburg). This was in compliance with the distance distribution functions, which also indicated aggregation for all proteins (Figure 48). In the case of VirGE57 the tail could also have been interpreted as the remaining residues of the VID, which unfolded and formed an elongated string. Also for VirGK34, the tail might have been interpreted as the amino acids, missing from the model. However, as all proteins showed signs of aggregation, it was more probable that the tail was an artifact in all cases.

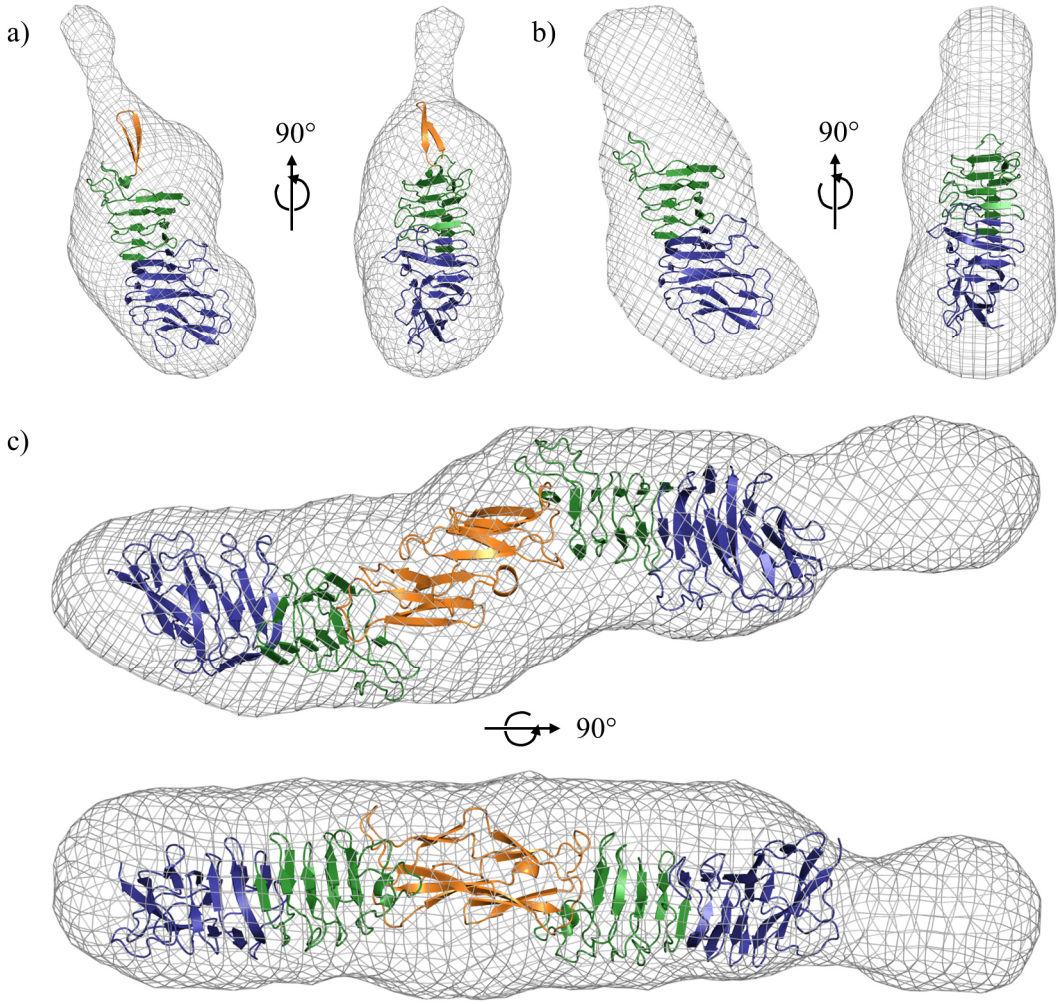


Figure 50: Envelopes calculated from the SAXS data and fitted structures or models for (a) VirGE57, (b) VirGK34 and (c) VirGE94.

4.1.4 New VirG-constructs

After the crystallization of VirGE94 the challenging task remained to identify soluble and crystallizable constructs of the more N-terminal parts of VirG, including the IcsB/ATG5 binding domain. As a first approach the boundaries of the VID were used as C-terminal boundaries for a number of new N-terminal constructs (VirGK9-K23). As there was no experimental data for convenient boundaries in the first 300 amino acids of VirG, different N-terminal boundaries were chosen based on secondary and tertiary structure predictions (Figure 51). In 1 l scale test-expressions none of these constructs showed expression, not in the soluble and neither in the insoluble fraction (Figure 52).

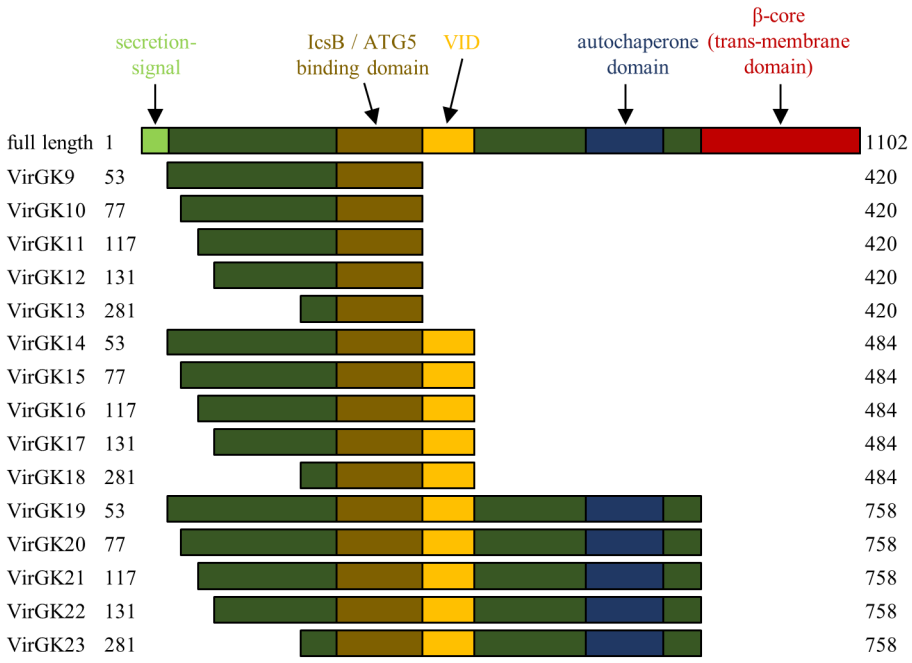


Figure 51: Overview of the VirG-constructs K9-K23.

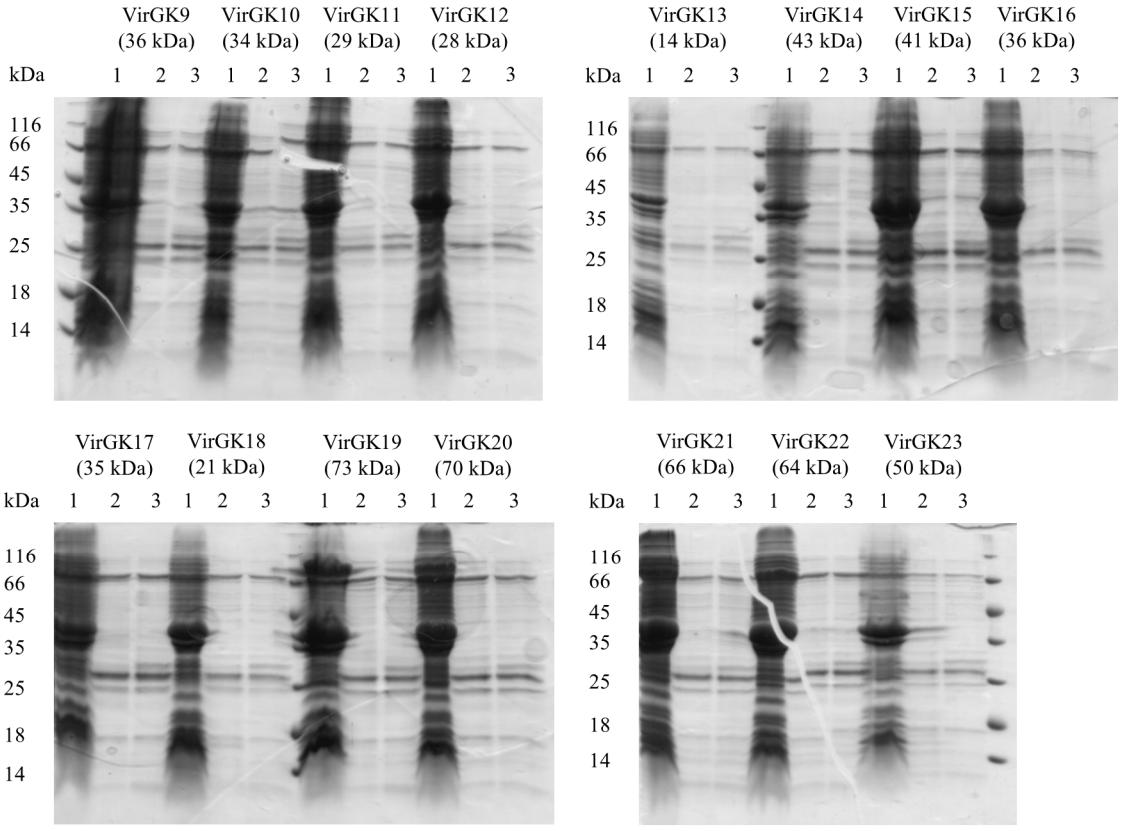
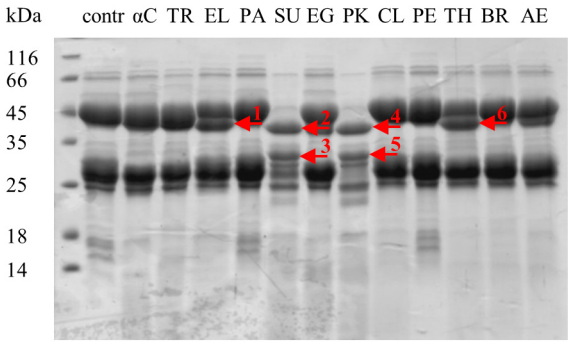


Figure 52: Coomassie stained SDS-gels for 1 l scale test-expressions of VirG-constructs K9-K23. 1) insoluble fraction, 2) soluble protein fraction (elution fraction from His-affinity column), 3) soluble protein cleaved by TEV-protease. No protein expression was detected.

The solubility of the constructs with N-termini between 53 and 281 was dramatically decreased. These results were in compliance with the results of the ESPRIT screening, where no soluble constructs with N-termini, preceding residue 340 were identified. Consequently in the next step it was tried to identify

soluble and crystallizable constructs with N-termini between the N-terminal boundaries of VirGE76 (340) and VirGE94 (419), which both had been expressed solubly before. Limited proteolysis of VirGE76 was performed to identify suitable N-termini (Figure 53). Six stable fragments were chosen and their N-termini were determined by Edman degradation. Of the identified N-termini, G415 and S418 were close to the N-terminus of the VID (N422), while S484 was close to its C-terminus (L483). Positions S349 and S383 were not located in the VID and thus were suitable N-termini for new constructs.



Fragment	N-terminus
1	S383
2	G415
3	S484
4	S418
5	S484
6	Q349

Figure 53: Coomassie stained SDS-gel for the limited proteolysis of VirGE76, using following restriction enzymes from the Hampton Research Proti-Ace and Proti-Ace2 Kits: αC α-chymotrypsin, TR trypsin, EL elastase, PA papain, SU subtilisin, EG endoproteinase Glu-C, PK proteinase K, CL clostripain, PE pepsin, TH thermolysin, BR bromelain, AE actinase. The table shows the determined N-termini of the marked bands.

New VirG-constructs were designed, using the two N-termini identified in the limited proteolysis experiment and three more N-termini, which were selected, using a rational approach, adding one predicted additional turn of the β-helix each to VirGE94 (VirGK28-K32). Additionally it was tested to remove the VID, to verify if the β-helix continued, upon its removal (VirGK33-K37). For this purpose several different linker length were tested to close the gap (VirGK24-K27) (Figure 54). The resulting 14 new VirG-constructs were subjected to 1 l scale test-expressions (Figure 55). For all constructs soluble protein was obtained. VirGK34 and K37 showed the strongest expression.

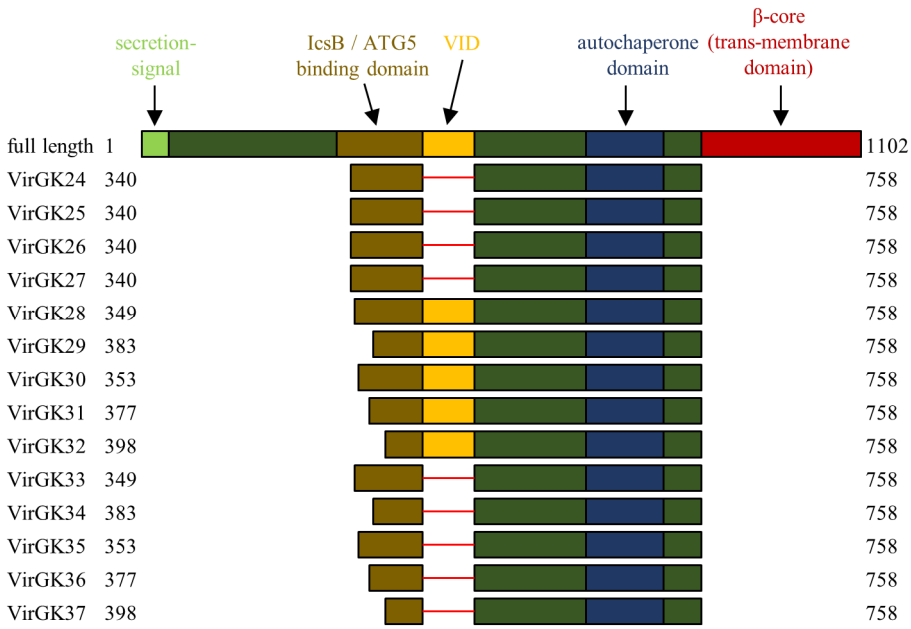


Figure 54: Overview of the VirG-constructs K24-K37.

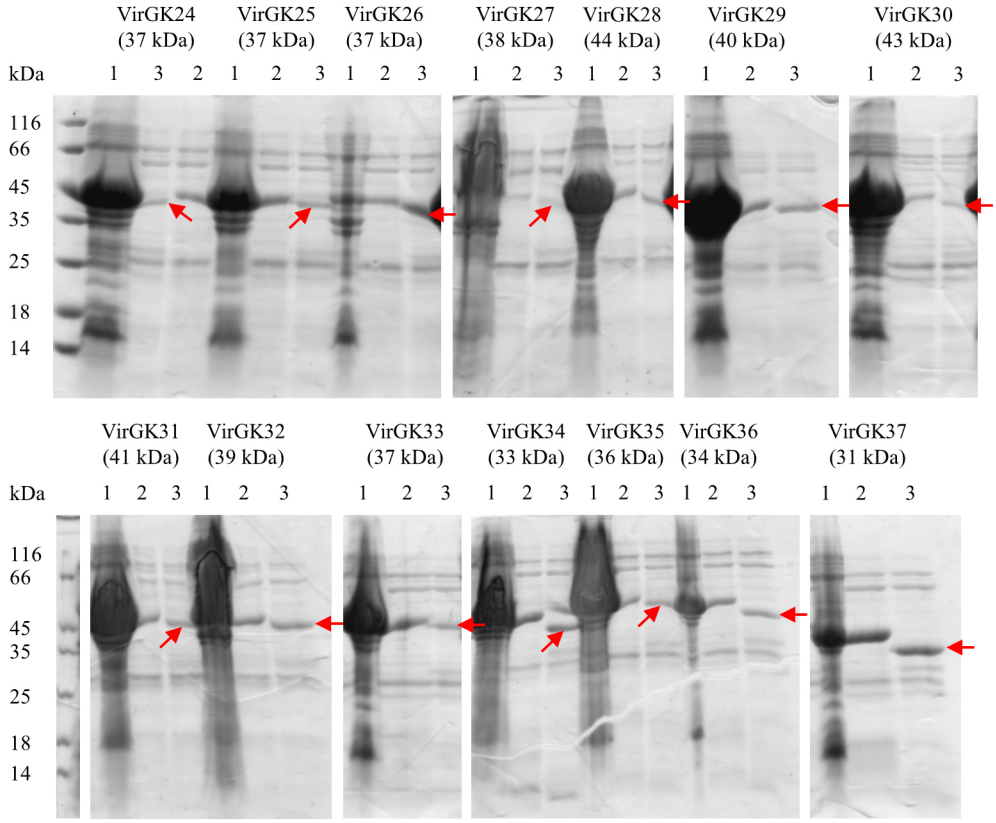


Figure 55: Coomassie stained SDS-gels for 1 l scale test-expressions of VirG-constructs K24-K37. 1) insoluble fraction, 2) soluble protein fraction (elution fraction from His-affinity column), 3) soluble protein cleaved by TEV-protease. Protein expression was detected for all constructs, most prominent expression for VirGK34 and K37. Red arrows mark VirG protein bands.

Subsequently the most promising of these constructs, namely VirGK25, K26 and K33-K37, were produced in 8 l large scale expressions. They were purified (Figure 56) and crystallization assays were set up. However, for none of these constructs initial crystallization conditions were identified.

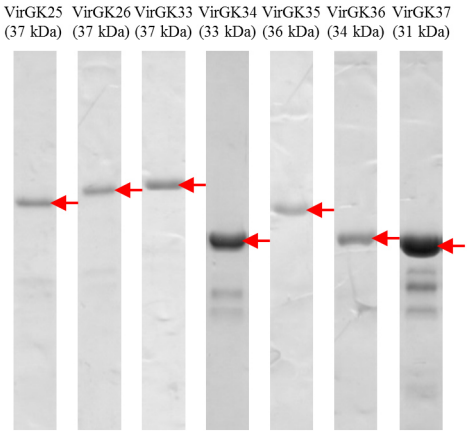


Figure 56: Coomassie stained SDS-gels showing samples from SEC for the VirG-constructs K25, K26, and K33-K37. All constructs were purified in varying purities and quantities. Red arrows mark VirG protein bands.

Finally a number of new constructs were designed, using the same N-terminal boundaries as before (340, 349, 383), but a truncated C-terminus. In the crystal structures of VirGE94 and of the autochaperone domain, the 18 C-terminal amino acids were flexible and not visible. These 18 residues were truncated in the new constructs (VirGK41-K43). Additionally the VID domain was deleted in the constructs VirGK38-K40 (Figure 57). 1 l scale test-expressions of the new constructs were performed (Figure 58). Soluble protein expression was comparable to the previous VirG-constructs, hence the 18

missing amino acids did not seem to influence protein expression, but they still might have had an influence on the crystallizability of these constructs. However these constructs were not expressed and purified in large scale in the timeframe of this thesis anymore. In the future it would be worth to express and purify at least VirGK40, the best expressing of the new constructs, to test for crystallization conditions.

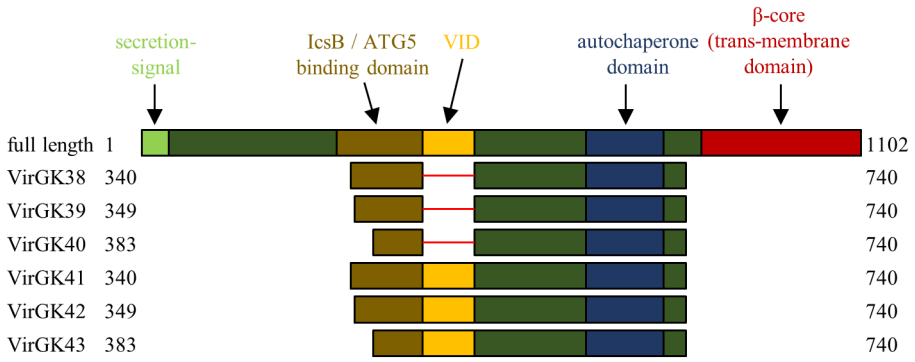


Figure 57: Overview of the VirG-constructs K38-K43.

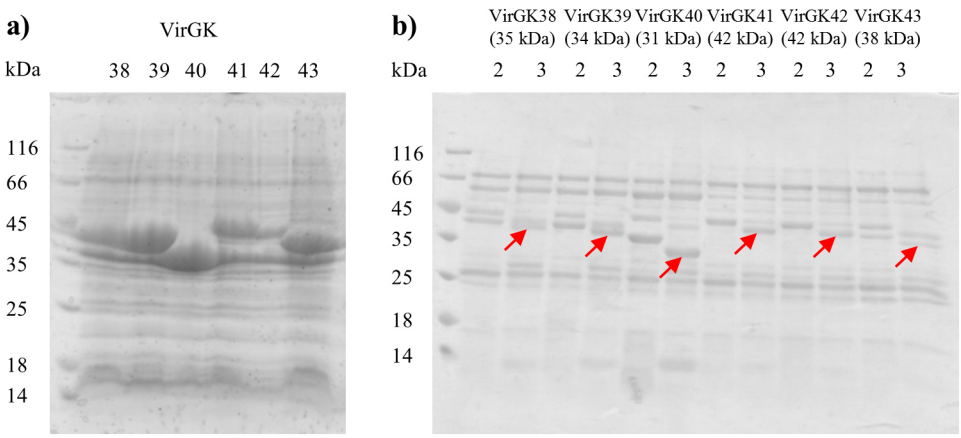


Figure 58: Coomassie stained SDS-gels for 1 l scale test-expressions of VirG-constructs K38-K43. a) SDS-gel of the insoluble fraction, b) SDS-gel of the: 2) soluble protein fraction (elution fraction from His-affinity column), 3) soluble protein cleaved by TEV-protease. Protein expression was detected for all of the constructs, most prominent expression for VirGK40. Red arrows mark VirG protein bands.

4.2 IcsB

4.2.1 Testing of the initial IcsB-constructs

The aim of this part of the work was the crystallization and the structural characterization of IcsB, which like for VirG required milligram amounts of soluble protein. Initially eleven truncated IcsB-constructs were designed, based on published domain boundaries and secondary structure predictions (IcsBK1-K11) (Figure 59). IcsBwt and the eleven truncated constructs were cloned as His-tagged proteins and coexpressed with IpgA in Rosetta 2 (DE3) cells. To identify suitable expression conditions, test-expressions were performed and all constructs except IcsBK4 were produced solubly and in high quantities (Figure 60).

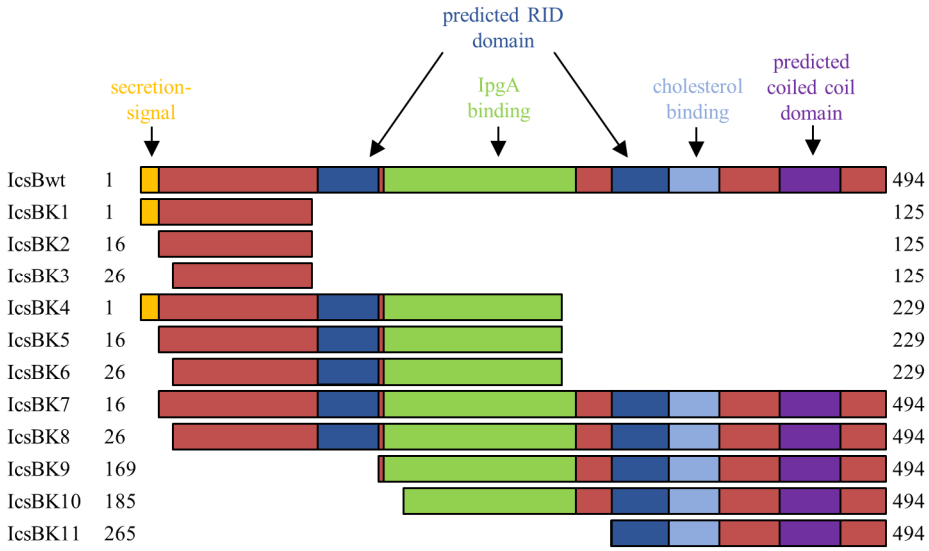


Figure 59: Overview of the IcsB-constructs K1-K11.

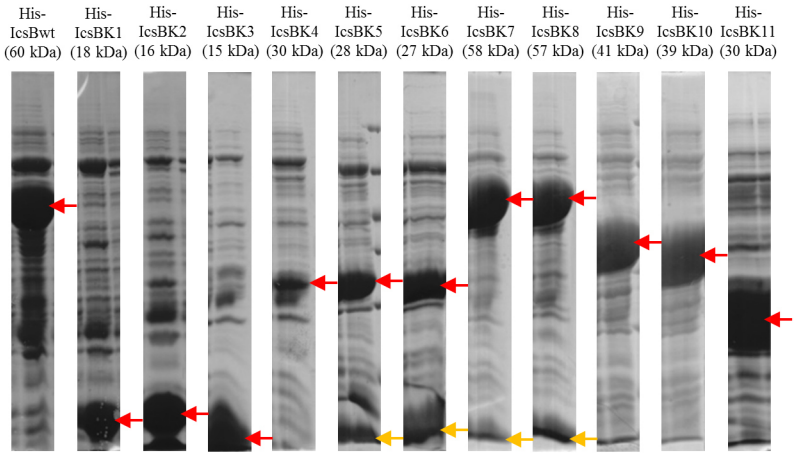


Figure 60: Coomassie stained SDS-gels for test-coexpressions of His-tagged IcsBwt/K1-K11 with IpgA. Only soluble protein fractions for the best expression condition are shown: TB, 20°C, 50 µM IPTG, overnight. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA. Strong expression for all of the constructs except for IcsBK4 was observed.

Most of the constructs were expressed and purified in large scale. Figure 61 shows the results of the purification of IcsBwt as an example. In Figure 61 b, a weak band with slightly lower MW than full length IcsB can be seen that coelutes in the same peak. This band was identified by tryptical digestion and mass spectrometry as mainly C-terminally truncated IcsB protein and is likely to be the result of premature termination of translation (Figure 66).

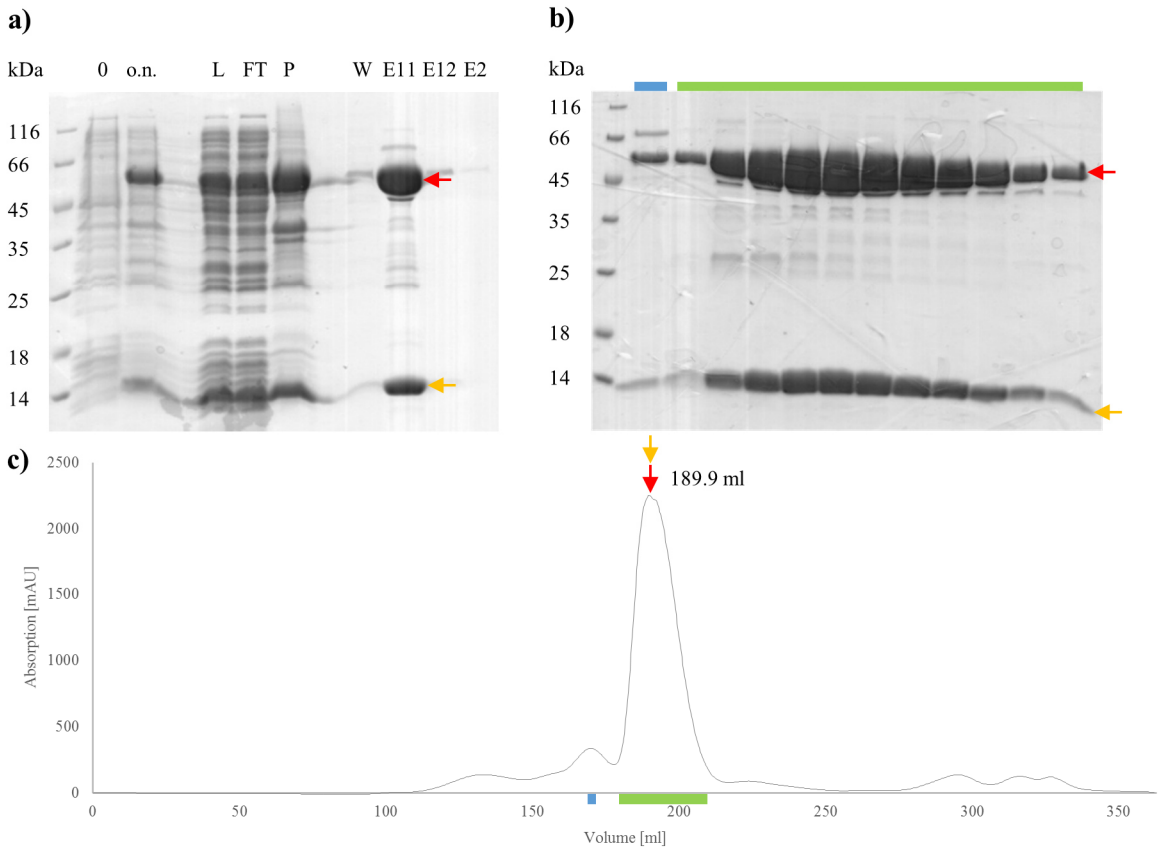


Figure 61: His-IcsBwt large scale expression and purification. As a first step the protein was purified, using a His-affinity column. In a second step the protein mixture was separated, using an S200 26/60 column. a) Coomassie stained SDS-gel of different steps in the His-affinity purification: 0, o.n.) samples taken directly from the culture before induction and after overnight expression, L) protein that was loaded to the column, FT) flow through of the column, P) pellet (insoluble protein), W) wash (30 mM imidazole), E11, E12) elution fractions (250 mM imidazole), E2) elution fraction (500 mM imidazole); b) Coomassie stained SDS-gel of different elution fractions from the S200 26/60 column. Elution fractions are color-coded and correspond to the fractions marked in c); c) Elution profile of the run on the S200 26/60 column. Elution volumes are indicated. Red arrows mark IcsBwt, orange arrows mark IpgA.

In Figure 62 and overview of the IcsB-constructs is given, which were produced and purified in large scale. In all cases high quantities of pure and soluble protein were obtained. During SourceQ IEC always a part of IpgA was separated from IcsB and could be purified separately. In contrast, the separated IcsB-constructs aggregated and could not be further used, indicating that IcsB was unstable without its chaperone IpgA. If IcsBwt was produced without coexpression of IpgA, soluble protein expression was lost. Hence IpgA was fundamental for the soluble expression and stabilization of IcsB.

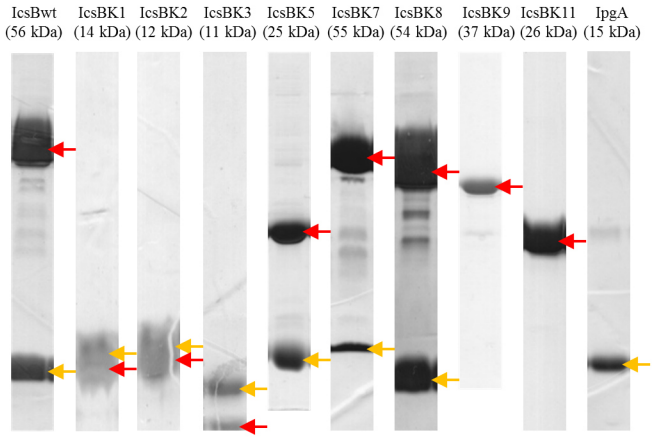


Figure 62: Coomassie stained SDS-gels showing samples from SEC for IcsBwt and the constructs K1-K3, K5, K7-K9, K11 and IpgA. All constructs were obtained in high purity and yield. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA.

Using the purified proteins, crystallization assays were set up, but no initial hits were obtained. In some cases crystalline material was observed, which always was identified as salt crystals.

4.2.2 Different rescue strategies for crystallization

As the initial crystallization screens didn't yield any results, a number of rescue strategies were applied. The first was the generation, expression and purification of a set of surface entropy reduction (SER) mutants of IcsBwt [189]. Using the online SER-server [190] six positions, suitable for SER, have been identified (IcsBSER1-6) (Figure 63). Test-expressions showed that all six constructs expressed as good as IcsBwt (Figure 64).

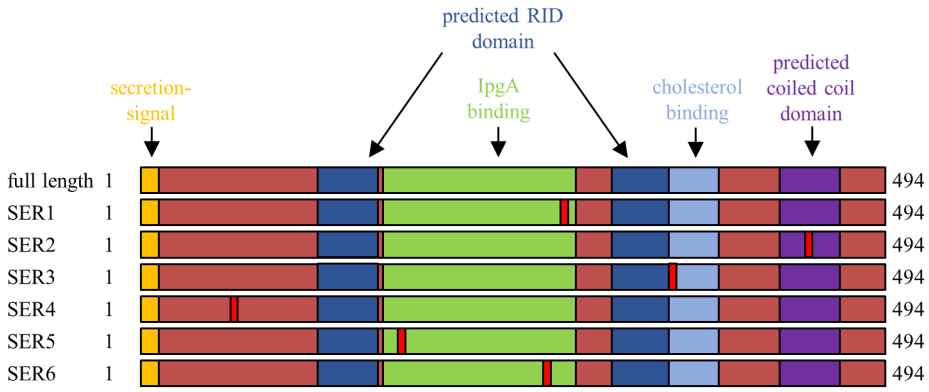


Figure 63: Overview of the IcsB-constructs SER1-6.

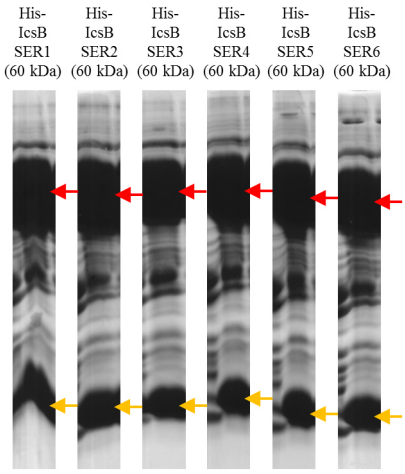


Figure 64: Coomassie stained SDS-gels for test-expressions of His-tagged IcsBSER1-6. Only soluble protein fractions for the best expression condition are shown: TB, 20°C, 50 µM IPTG, overnight. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA. Strong expression was observed for all constructs.

All SER-mutants were expressed and purified in large scale (Figure 65). Crystallization assays were set up, but no initial hits were obtained.

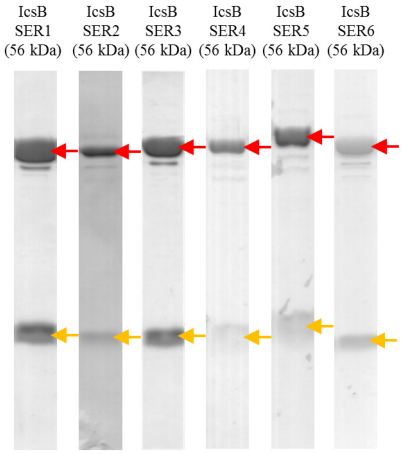


Figure 65: Coomassie stained SDS-gels, showing samples from SEC for IcsBSER1-6. All constructs were obtained in high purity. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA.

As a second rescue strategy the methylation of surface lysines was tested. This approach was tested using IcsBK3. Only small amounts of protein precipitated during the methylation reaction and the final yields were high with 79%. Crystallization assays were set up, using the methylated protein, but also this approach didn't result in any initial hits.

In situ proteolysis [191] was tested as another means to crystallize IcsBwt. Trypsin, chymotrypsin, elastase, endoproteinase Glu-C and clostripain, in total five proteases, some of which produced stable fragments in limited proteolysis experiments (compare Figure 66), were used at concentrations of 1:1000 to 1:10000. However, also this approach didn't yield any initial hits.

In another experiment, YCl_3 was supplemented to several commercially available screens at a concentration of 100 µM. Yttrium is a transition metal with atomic number 39, hence its trivalent ion can provide orbitals for complex formation in different geometries. In these experiments individual IpgA or complexes of IpgA with small IcsB-constructs were used. IpgA had a low pI, meaning it was negatively charged at the pH-values, usually present in commercial screens. Hence the positively charged yttrium ions could have mediated crystal contacts between different copies of IpgA. Indeed the

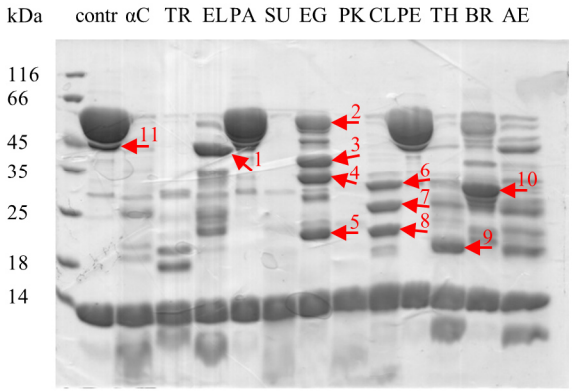
growth of needle shaped crystals was observed over the course of several days to weeks, which were identified as salt crystals, in this case most probably yttrium salt crystals.

The combination of all IcsB-constructs with all rescue strategies would have resulted in a number of experiments that one person wouldn't have been able to handle anymore. For this reason, in a next step, new IcsB-constructs were designed, using boundaries based on experimental data.

4.2.3 Design of new IcsB-constructs

Limited proteolysis experiments were performed using IcsBwt, to identify suitable boundaries for the design of new IcsB-constructs. These experiments were done, using the Proti-Ace and Proti-Ace2 Kits of Hampton Research. Samples of the digested protein were analyzed on an SDS-gel and the results are shown in Figure 66. Several proteases produced stable fragments of IcsBwt, especially elastase (EL), endoproteinase Glu-C (EG-C), clostripain (CL), thermolysin (TH) and bromelain (BR). The N-terminal boundaries of the fragments were determined by Edman degradation and their MWs by mass spectrometry. For the fragments 3, 7 and 9 several different N-termini were identified, which meant that the band was composed of different fragments. Fragments 2 and 5 had an N-terminus before the start of the actual protein inside the His-tag or the linker. The C-terminal boundaries were calculated, using the N-termini and the MWs of the fragments. For fragment 9 the MW could not be determined, due to experimental problems with the sample. In all other cases the calculated C-termini and the according calculated MWs of the fragments are indicated. In most cases the experimental and the calculated MWs were similar. Interestingly, in the cases where several N-termini were determined, only the longest of the fragments yielded a plausible C-terminus, as if one of the other N-termini would have been used, the C-terminus of the fragment would have exceeded the full length protein, which was not possible. In these cases a 'too big' (tb) is indicated in the table in Figure 66 b.

a)



b)

Fragment	N-terminus	Exp. MW [Da]	C-terminus (Theor. MW [Da])
1	S6	44967	I399 (44950)
2	N-12	53428	K457 (53440)
3	I177/K183	36347	N493 (36340)/tb
4	V214	32131	I494 (32140)
5	N-12	21192	E176 (21250)
6	A235	29529	N493 (29480)
7	Q293/T333/ T393	22987	N493 (22960)/tb/tb
8	M1	20863	K183 (20800)
9	I230/I9		
10	N236	29861	I494 (29520)
11	F8	49055	N438 (49060)

Figure 66: a) Coomassie stained SDS-gel for the limited proteolysis of IcsBwt, using following restriction enzymes from the Hampton Research Proti-Ace and Proti-Ace2 Kits: contr) control without enzyme, αC) α-chymotrypsin, TR) trypsin, EL) elastase, PA) papain, SU) subtilisin, EG) endoproteinase Glu-C, PK) proteinase K, CL) clostripain, PE) pepsin, TH) thermolysin, BR) bromelain, AE) actinase. b) N-termini, experimental MWs and calculated C-termini with corresponding theoretical MWs of the bands marked in a). See text for a more detailed explanation.

To gain insight into the substructure of IcsB, aliquots of the full length protein were digested with elastase, endoproteinase Glu-C, clostripain, thermolysin and bromelain for one hour at 37°C. Afterwards the reaction mixture was separated on an S200 10/30 column. In all cases only a single peak eluted from the column at a volume of 13.9 ml. When it was analyzed on an SDS-gel, a number of protein fragments were identified in the according fractions. Hence IcsB maintained its tertiary structure, even when it was cut into several pieces. IcsB seemed to be a compact protein, exposing several loops to the solvent that were cut, but didn't seem to contain loosely attached subdomains, which would have been separated from the protein upon digestion.

The boundaries, identified in the limited proteolysis experiment, were used to design new IcsB-constructs (IcsBK12-K29, K32) (Figure 68). Test-expressions showed a widely varying amount of solubly expressed protein (Figure 68). Especially the constructs with C-terminal boundaries between 391 and 464 didn't express solubly. Maybe in these constructs some C-terminal domain that plays a role in the stability of IcsB, was disrupted.

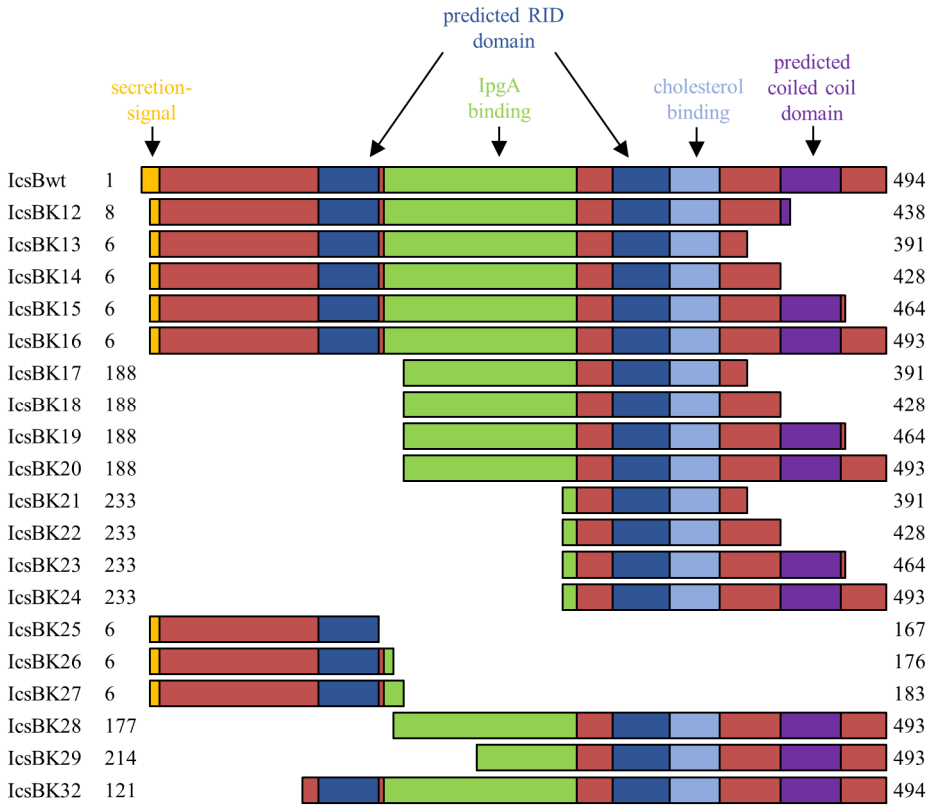


Figure 67: Overview of the IcsB-constructs K12-K29 and K32.

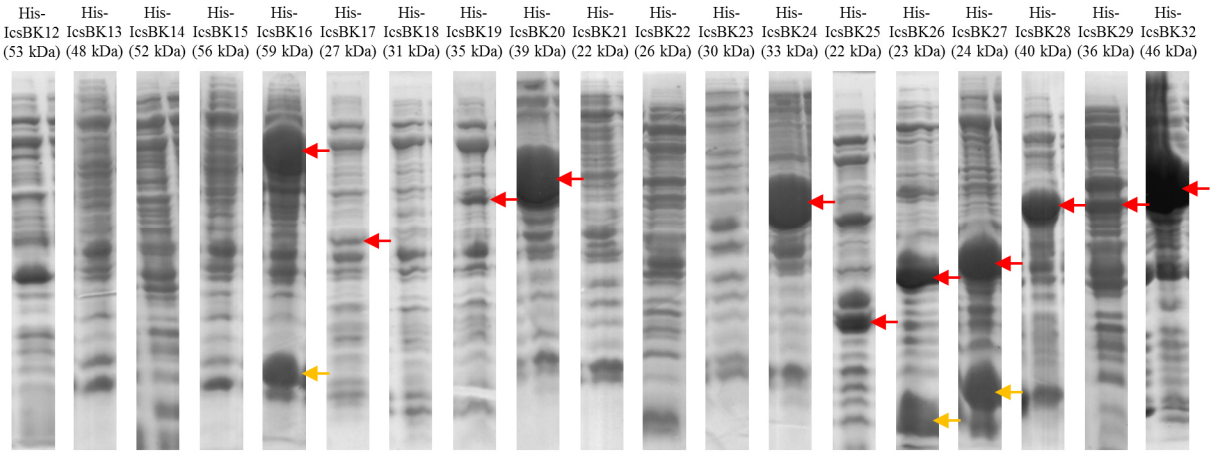


Figure 68: Coomassie stained SDS-gels for test-expressions of IcsB-constructs K12-K29 and K32. Only soluble protein fractions for the best expression condition are shown: TB, 20°C, 50 µM IPTG, overnight. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA. Strong expression for IcsBK16, K20, K24-K29, K32, weak expression for K17, K19 and no expression for the rest of the constructs was observed.

Of the constructs that showed soluble protein expression, K16, K24, K26, K27, K29 and K32 were expressed and purified in large-scale (Figure 69). Crystallization assays were set up, but no initial crystallization conditions were identified.

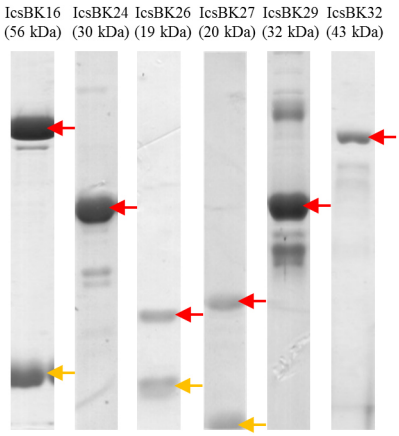


Figure 69: Coomassie stained SDS-gels, showing samples from SEC for IcsBK16, K24, K26, K27, K29 and K32. All constructs were obtained in high purity. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA.

In a new approach the reconstitution of nearly full length IcsB from two protein fragments was tested. The region in between residues 196-230 was predicted as unfolded by DISOPRED2 and in limited proteolysis experiments a number of proteases cut in this region (176-236), indicating that it was flexible and accessible from the solvent. By limited proteolysis and SEC it was shown that IcsB maintained its tertiary structure upon digestion, indicating that fragments of the protein have an affinity for each other. Hence it was tried to remove the unfolded region from the protein, using two purified IcsB-constructs, the N-terminal fragment IcsBK3 (26-125) and the C-terminal fragment K24 (233-493), which should bind to each other to reconstitute nearly full length IcsB. The two proteins were mixed, incubated and separated on an S200 10/30 column. However, in contrast to previous results, the two fragments didn't interact with each other. They eluted in two separate peaks, corresponding to the elution volumes of the individual proteins, and no additional peak, corresponding to the complex, appeared.

In summary, 36 truncated or modified IcsB-constructs were tested and combined with different rescue strategies for crystallization, not leading to any initial crystallization condition. At this point further attempts to identify a crystallizable construct of IcsB were stopped and focus was lain on the biophysical characterization and interaction studies.

4.2.4 Biophysical characterization of IcsB

A thermofluor-assay was performed, using IcsBwt (in complex with IpgA) (Figure 70). The best buffer conditions for the complex were HEPES/TRIS pH 7.0-8.0 with ≥200 mM NaCl. In these conditions the complex had a melting temperature of around 50°C (Figure 70 b). At pH values below 5.0 the complex didn't have distinct melting curves anymore and was probably denaturated. No stabilizing additives, except glycerol at concentrations from 5-20%, were identified.

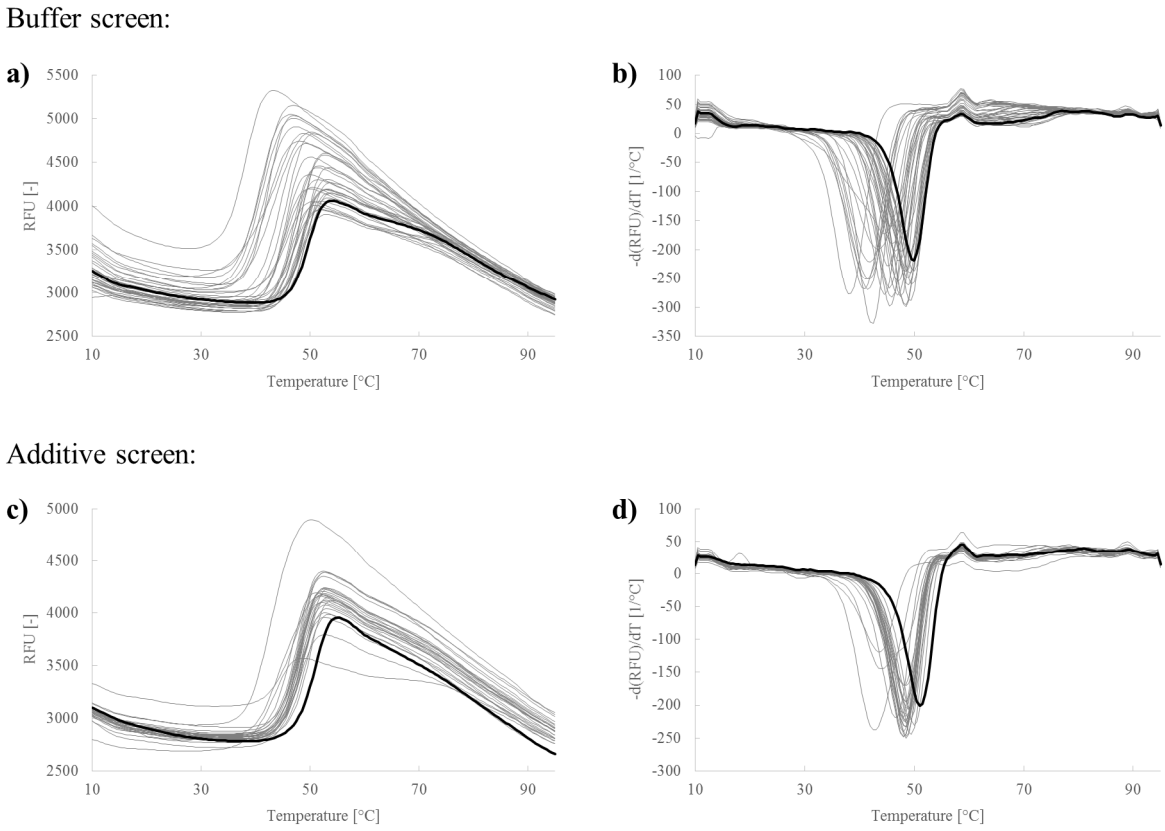


Figure 70: Thermofluor-assay for IcsBwt (in complex with IpgA). a) Melting curves and b) melting peaks of a buffer screen. c) Melting curves and d) melting peaks of an additive screen. The complex had a melting temperature of about 50°C.

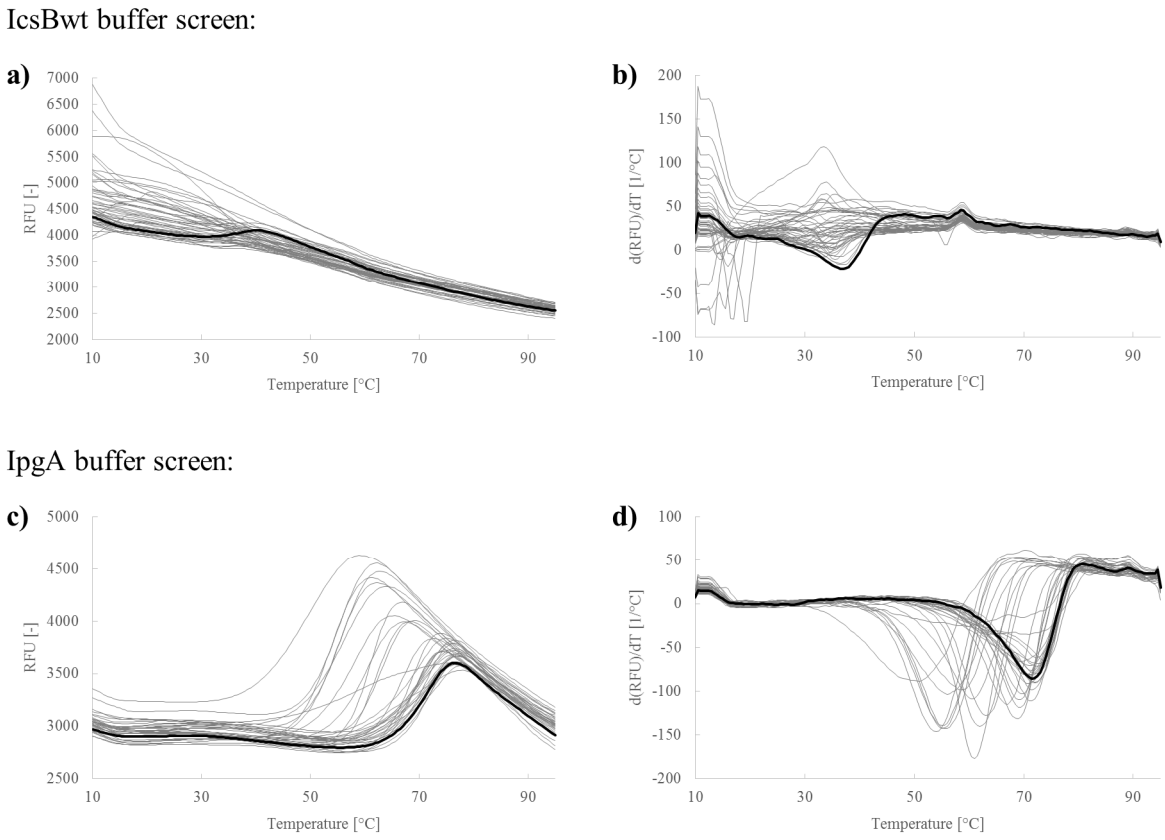


Figure 71: Thermofluor-assay for individual IcsBwt and IpgA. a) Melting curves and b) melting peaks of IcsBwt in a buffer screen. The highest melting temperature is around 37°C. c) Melting curves and d) melting peaks of IpgA in a buffer screen. The highest melting temperature is around 72°C.

As has been described before, it was possible to separate the IcsB-IpgA complex, by SourceQ IEC. Individual IpgA remained soluble, while IcsB aggregated and ran in the void of an S200 16/60 column. The separated IcsB and IpgA were used for thermofluor-assays (Figure 71). Individual IcsB showed indistinct melting curves with only a minor maximum, indicating that it was mostly denatured already in the beginning. Still, a melting temperature of around 37°C was determined. Single IpgA had a melting temperature of around 72°C. The best buffer conditions were 50 mM HEPES pH 7.0 without NaCl for IcsB and 50 mM TRIS pH 8.0, 100 mM NaCl for IpgA.

To test whether IcsBwt (in complex with IpgA) was properly folded, CD-spectra of the complex, as well as individual IpgA and IcsBK9 (169-494), a soluble C-terminal IcsB-construct that didn't form a complex with IpgA, were recorded (Figure 72). The resulting ratios of secondary structure elements were not untypical, compared to other proteins [192].

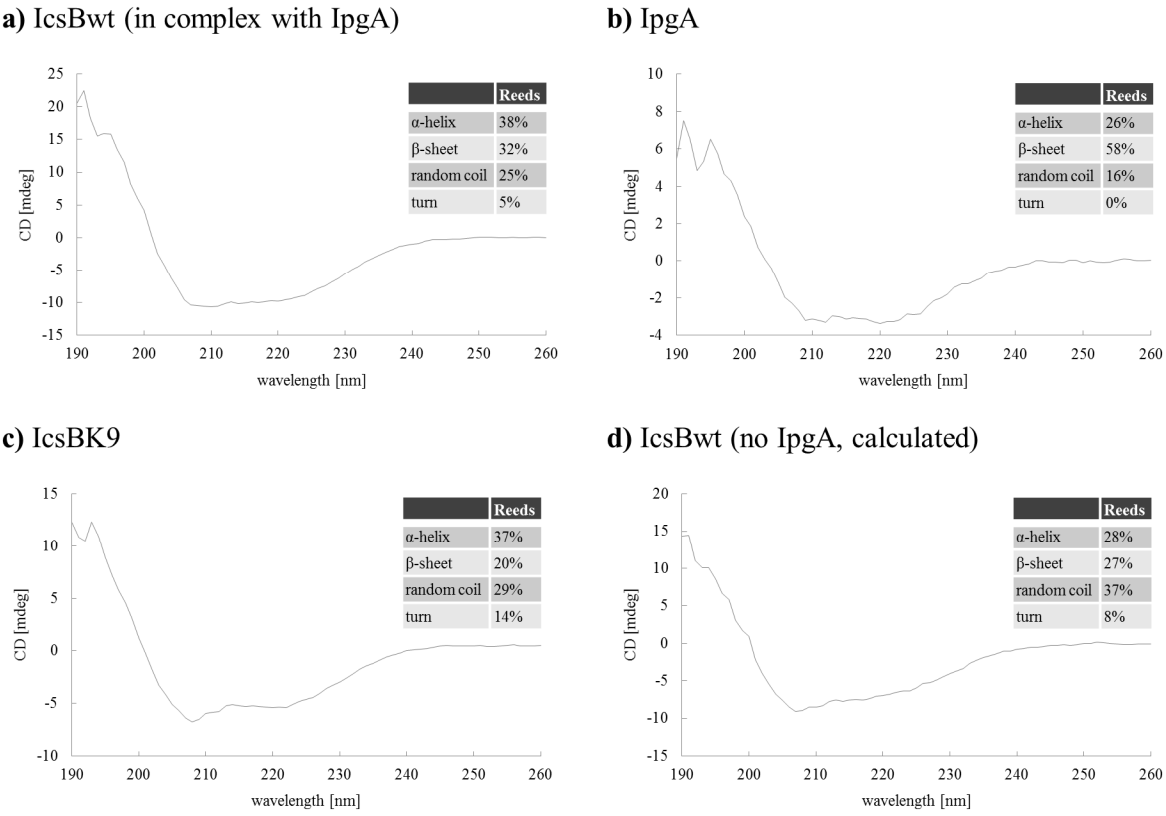


Figure 72: CD-spectra of the four proteins or complexes: a) IcsBwt (in complex with IpgA), b) IpgA, c) IcsBK9 and d) IcsBwt (no IpgA). The spectra of the first three proteins IcsBwt (in complex with IpgA), IpgA and IcsBK9 were experimental CD-spectra. The CD-spectrum of IcsBwt (no IpgA) was calculated by the subtraction of the two CD-spectra of IcsBwt (in complex with IpgA) and IpgA. Reeds method was used to calculate the ratios of secondary structure elements [193].

4.2.5 Mapping of the IpgA binding site

In [131] it was shown that the smallest IpgA binding fragment of IcsB, spanned the residues 171-247. However in purifications of small N-terminal constructs of IcsB, comprising the first 125 residues, IpgA repeatedly copurified (Figure 62). This led to the hypothesis that the IcsB-CBD was actually located in the N-terminal part of IcsB. For this reason a number of IcsB-constructs, all coexpressed with IpgA, were purified and tested for the presence of IpgA in a systematic manner (Figure 73). N-terminal truncation of IcsB resulted in a loss of interaction with IpgA, even if only the first 168 amino acids were deleted (IcsBK9). On the other hand IpgA still copurified with all the C-terminally truncated IcsB-constructs. The shortest construct of IcsB that copurified IpgA was IcsBK3 (26-125).

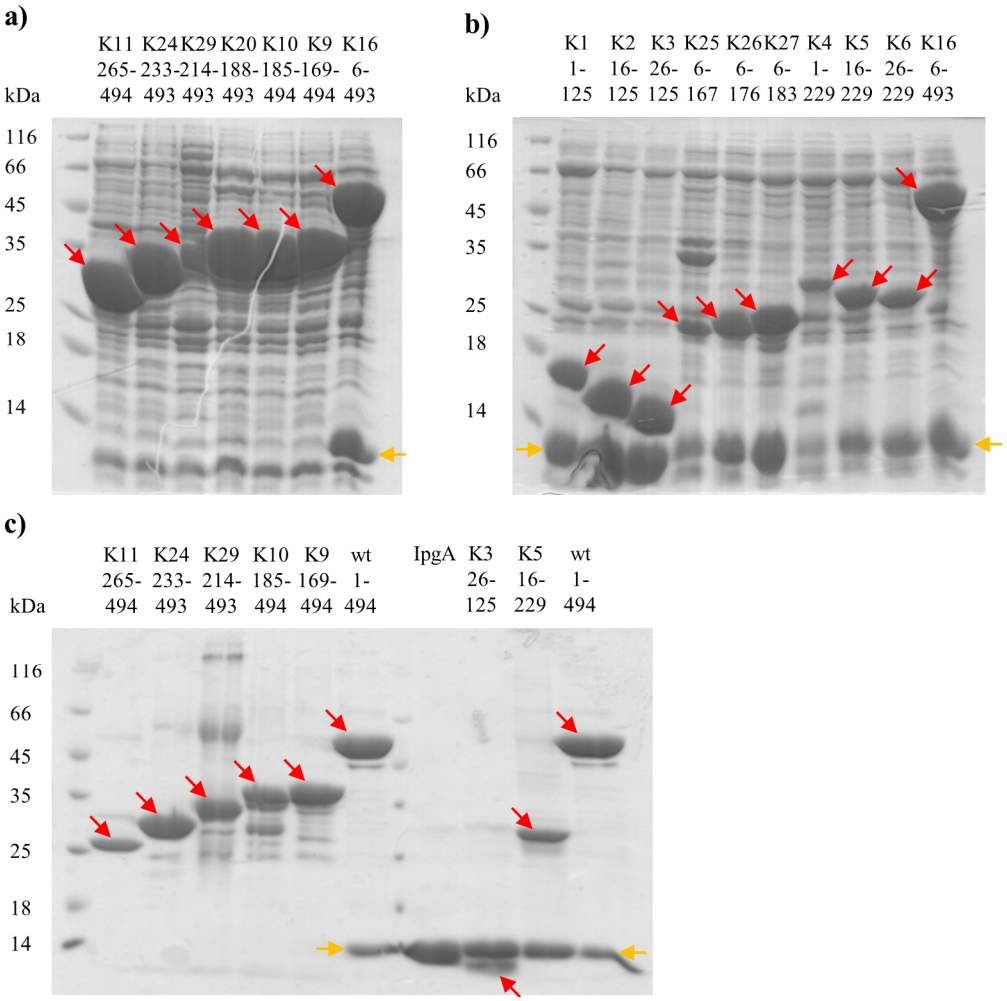


Figure 73: Coomassie stained SDS-gels for purifications of different His-tagged IcsB-constructs coexpressed with untagged IpgA. a) N-terminal truncations of IcsB, b) C-terminal truncations of IcsB with the indicated boundaries. c) A part of the IcsB-constructs after SEC purification. Red arrows mark the corresponding IcsB-constructs, yellow arrows mark IpgA. IpgA copurifies with none of the N-terminally truncated constructs, except for IcsBK16, which only lacks the first five amino acids, but with all of the C-terminally truncated constructs.

SAXS experiments were performed, using IcsBK3 (in complex with IpgA) and individual IpgA (Figure 74). The distance distribution functions of both proteins were nearly unimodal, indicating that the proteins in solution were homogenous and didn't form aggregates. The SAXS data were used to calculate the MW of both constructs, shown in Table 18. The MW of IpgA in solution corresponded to a dimer, while the MW of IcsBK3 corresponded to one molecule of IcsBK3, binding an IpgA dimer.

Table 18: MWs and the corresponding oligomerization states of IpgA and IcsBK3 (in complex with IpgA), calculated from the SAXS data.

Construct	Molecular weight [kDa]	Oligomerization state
IpgA	34.4	2xIpgA
IcsBK3-IpgA	45.2	IcsBK3+2xIpgA

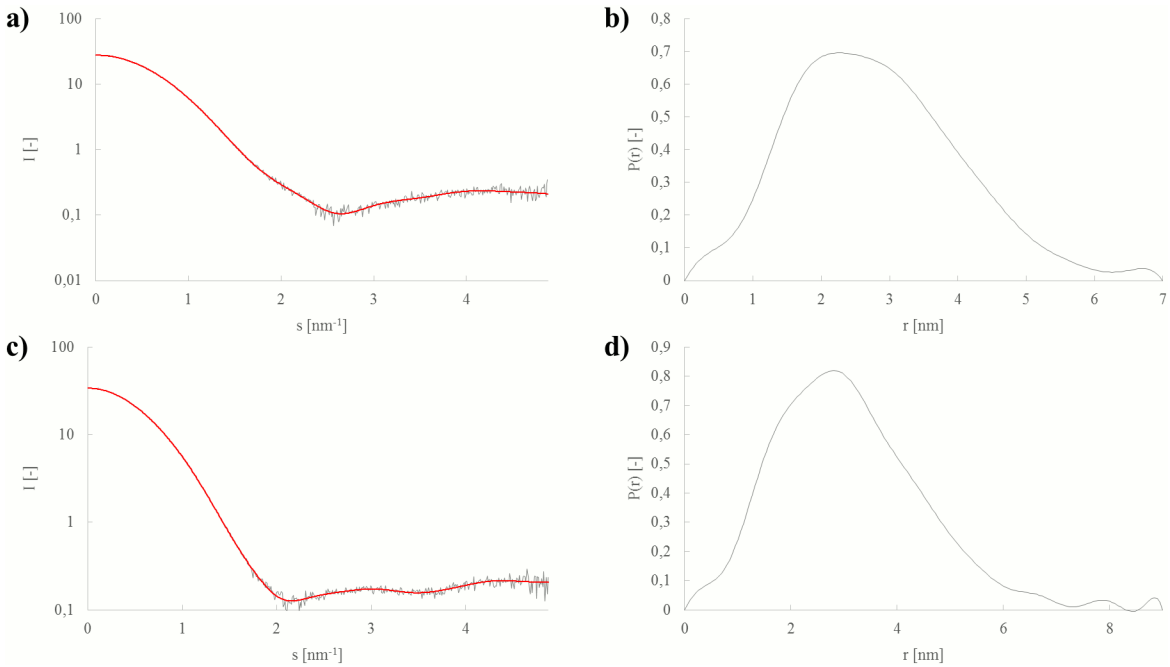
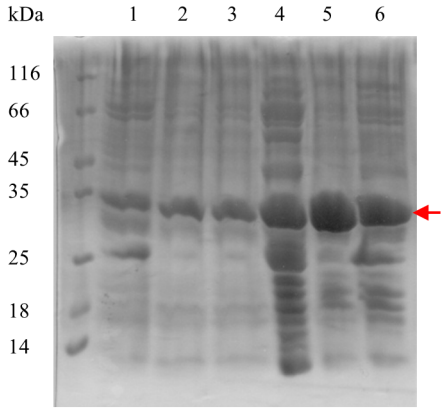


Figure 74: (a), (b) IpgA and (c), (d) IcsBK3 (in complex with IpgA) have been used in SAXS experiments. In the left column experimental (grey) and regularized (red) intensities are plotted against the scattering length s. In the right column particle distance distribution functions P(r) are shown.

4.3 ATG5

4.3.1 Purification and characterization of human ATG5

Like for VirG and IcsB, the objective for ATG5 in this part of the work, was the crystallization and structural characterization, which required milligram amounts of protein. Initially only one human ATG5-construct was purified, the His-tagged full length protein. Test-expressions were performed (Figure 75) and the best expression condition was identified as expression in TB at 20°C overnight and induction with 200 µM IPTG.



Sample	Temperature [°C]	IPTG [µM]	Incubation time
1	25	/	3 h
2	20	200	3 h
3	25	200	3 h
4	25	/	o.n.
5	20	200	o.n.
6	25	200	o.n.

Figure 75: Coomassie stained SDS-gel for test-expressions of His-tagged ATG5. Only soluble protein fractions are shown. The best expression condition was: TB, 20°C, 200 µM IPTG, overnight. The red arrow marks ATG5.

ATG5 was expressed and purified in large scale (Figure 76). About half of the protein was expressed in the insoluble fraction, while the other half expressed solubly. If the soluble protein was kept on ice for too long or if it was concentrated too high, heavy precipitation occurred.

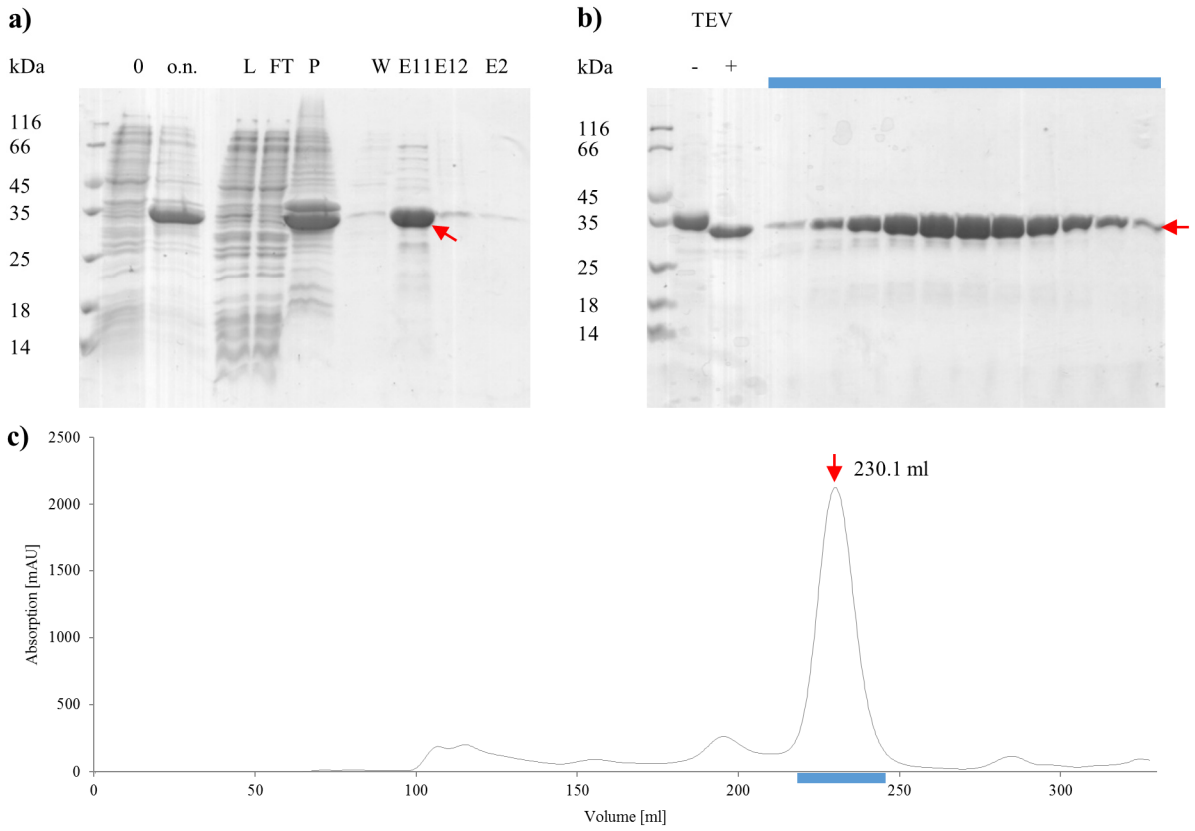


Figure 76: ATG5 large scale expression and purification. First the protein was purified, using a His-affinity column, in a second step by SEC, using an S200 26/60 column. a) Coomassie stained SDS-gel of different steps in the His-affinity purification: 0, o.n.) samples taken directly from the culture before induction and after overnight expression, L) protein that was loaded to the column, FT) flow through of the column, P) pellet (insoluble protein); W) wash, E11, E12) elution fractions (250 mM imidazole), E2) elution fraction (500 mM imidazole); b) Coomassie stained SDS-gel of different elution fractions from the S200 26/60 column. Fractions are color-coded and correspond to the fractions marked in c). -TEV) cleavage control without protease, +TEV) cleaved protein; c) Elution profile of the run on the S200 26/60 column. Elution volumes are indicated. Red arrows mark ATG5.

As the protein easily precipitated, a thermofluor-assay was performed to search for stabilizing buffer conditions (Figure 77). The best identified buffer conditions were 50 mM HEPES/TRIS pH 7.0-8.0, 500 mM NaCl. Addition of 5-20% glycerol further stabilized ATG5. Finally a buffer of 50 mM HEPES pH 7.0, 500 mM NaCl and 5% glycerol was identified as the ideal buffer for the purification of ATG5. Using this buffer, precipitation of ATG5 was minimized.

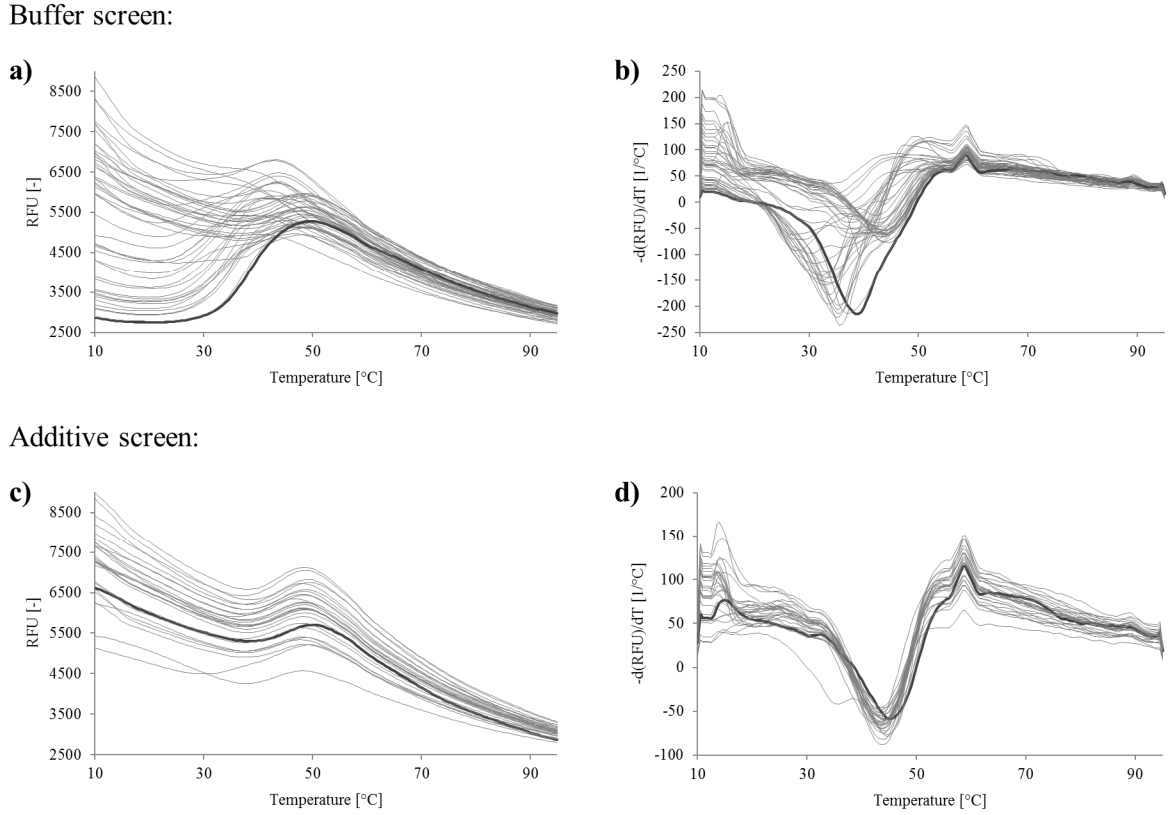


Figure 77: Thermofluor-assay for ATG5. a) Melting curves and b) melting peaks of a buffer screen. c) Melting curves and d) melting peaks of an additive screen. ATG5 had a melting temperature of about 39°C in the buffer screen. Addition of 5-20% glycerol stabilized it up to 45°C.

4.3.2 ATG5 crystallization and structure

Crystallization setups were done for ATG5 and after several days very thin, needle shaped crystals were identified in condition 93 of the PEGs suite (0.2 M Li_3citrate, 20% PEG 3350, 11 mg/ml, 4°C) (Figure 78). The crystals in this condition were reproduced and optimized, using custom 96-well grid screens around this condition. Finally larger crystals were obtained in 24-well hanging drop setups (0.35 M Li_3citrate, 32.2% PEG 3350, 10 mg/ml, 20°C). The crystals in the optimized condition were still needle shaped and mostly extremely thin, but several reached dimensions of up to 500x30x30 µm, which were suitable for data collection.

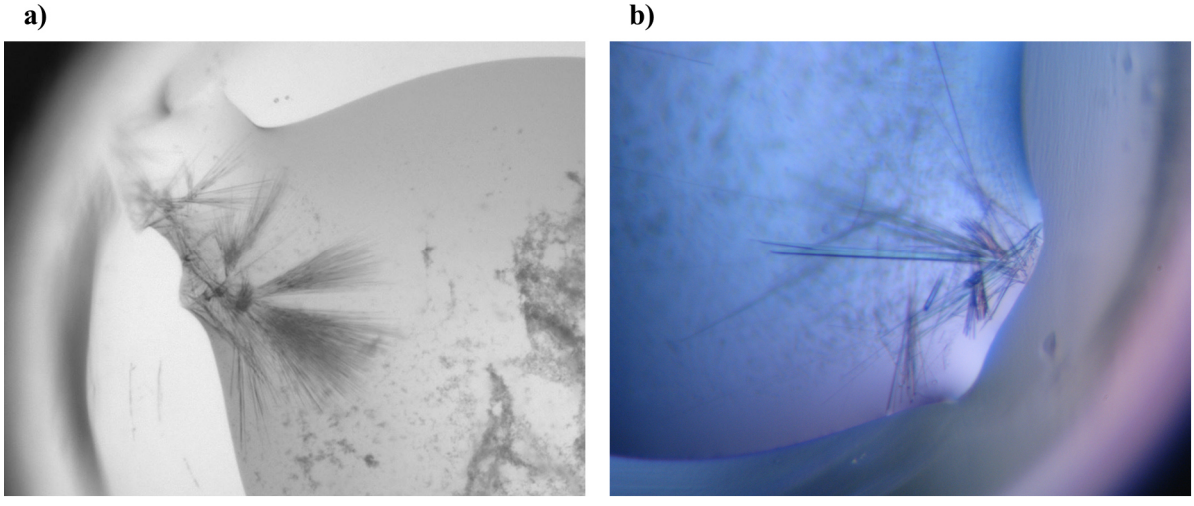


Figure 78: ATG5 crystals. a) Condition 93 from PEGs suite (0.2 M Li_3Citrate, 20% PEG 3350, 11 mg/ml, 4°C); b) reproduced in a 96-well grid-screen (0.325 M Li_3Citrate, 24% PEG 3350, 10 mg/ml, 20°C).

The crystals diffracted to a resolution of around 4 Å in house. A dataset was collected, using the Pilatus 2M detector at beamline PXIII of the SLS synchrotron at a wavelength of 1.000 Å and a temperature of 100 K, where they diffracted to a resolution of around 3.1 Å. ATG5 crystallized in the $P4_12_12$ space group with one molecule in the asymmetric unit. The structure was solved by molecular replacement, using ATG5 of the ATG5-12/16 complex from PDB-entry 4GDK as a search model. The structure of ATG5 is shown in Figure 79.

Figure 79: Crystal structure of ATG5.

Table 19: Data collection and refinement statistics for ATG5.

Data collection	
X-ray source	PXIII, SLS
Wavelength [Å]	1.000
Resolution [Å]	3.10 (3.21 – 3.10)
Space group	$P4_12_12$
Unit-cell parameters [Å]	75.0, 75.0, 115.3 90.0, 90.0, 90.0
R_{merge}	0.209 (0.801)
R_{meas}	0.214 (0.817)
CC1/2	0.998 (0.970)
‹I/σ(I)›	24.27 (6.27)
Completeness [%]	100.0 (100.0)
No. of reflections	160665 (16453)
Multiplicity	25.1 (26.5)
Refinement	
Resolution [Å]	53.0 – 3.10
No. of reflections	6398 (620)
R_{work}	0.190 (0.265)
R_{free}	0.218 (0.327)
Asymmetric unit	1 molecule
No. of atoms	
Protein	2189
Ligands	0
Water	0
Average *B* factor [Å^2]	
Protein	33.1
Ligands	/
Water	/
R.m.s. deviations	
Bond length [Å]	0.008
Bond angles [°]	1.29
Ramachandran favored [%]	98
Ramachandran outliers [%]	0.0

ATG5 was used for SAXS experiments to determine its oligomerization state in solution (Figure 80 a, b). The distance distribution function was unimodal, indicating that the protein was homogenous in solution and didn't form aggregates. Fitting of the calculated scattering curves of an ATG5 monomer to the experimental SAXS data, resulted in a good fit with a χ^2-value of 2.8 (Figure 80 c), indicating that ATG5 was monomeric in solution.

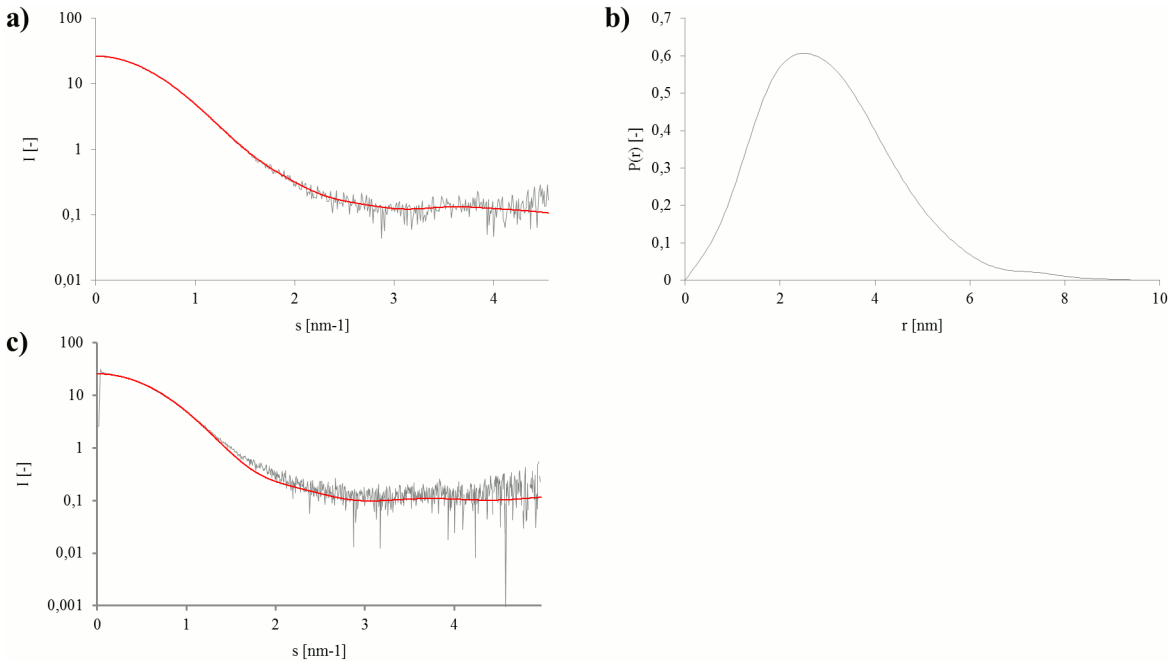


Figure 80: ATG5 has been used in SAXS experiments. a) Experimental (grey) and regularized (red) intensities are plotted against the scattering length s. b) Particle distance distribution function P(r). c) Fitting of the calculated SAXS scattering curve of an ATG5 monomer (red) to the experimental scattering curve (grey). A good fit with a χ^2-value of 2.8 was obtained.

4.3.3 Function of the phosphorylation of threonine 75

ATG5 residue T75 can be phosphorylated by the p38 MAP-Kinase, resulting in the inhibition of autophagy, by preventing the fusion of autophagosomes and lysosomes [162]. However, the mechanism of the inhibition has not been identified yet. One possible explanation might be that, by the phosphorylation of T75, the interaction of ATG5 with ATG16L1 or TECPR-1 is influenced, as they play major roles in the maturation of autophagosomes and their fusion with lysosomes [165]. A possible way to change the affinity of ATG5 to ATG16L1 or TECPR-1, might be a change in the tertiary structure of ATG5 upon phosphorylation of T75.

To investigate the role of the phosphorylation of T75, two ATG5 mutants were prepared, T75E and T75A (Figure 81). Replacing threonine by glutamic acid results in a phosphomimetic mutant. The size and the charge of glutamic acid mimic a phosphorylated threonine. The alanine mutant served as a negative control, as alanine is uncharged and cannot be phosphorylated. Therefore the structure of the T75A mutant should resemble the structure of unphosphorylated ATG5 under all conditions. If the structures of both mutant forms of ATG5 were solved, an influence of the phosphomimetic mutation on the tertiary structure of ATG5 would have been revealed. Test-expressions were performed for the two mutant forms of ATG5 and both proteins expressed as good as or better than ATG5wt (Figure 82). Subsequently the two mutant forms were expressed and purified in large scale.

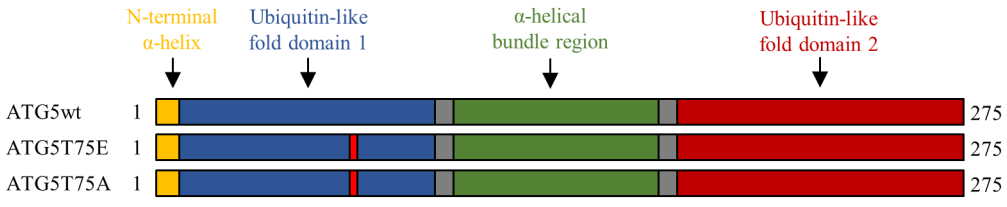


Figure 81: Overview of the ATG5-constructs T75E and T75A.

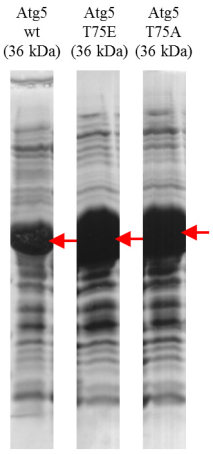


Figure 82: Coomassie stained SDS-gels for test-expressions of ATG5wt, T75E and T75A. Only the soluble protein expression of the best expression condition is shown: TB, 50 µM IPTG, 20°C, overnight. Red arrows mark the corresponding ATG5-constructs. The expression of the two mutant forms was stronger than the expression of ATG5wt.

4.3.3.1 Crystallization and structure of ATG5T75E and T75A

A completely new screening for crystallization conditions was performed for ATG5T75E and T75A, to exclude the possibility that, by using the same crystallization condition as for the ATG5wt, a certain conformation was imposed on the mutant forms. However, the only crystallization condition that yielded crystals was the one where already ATG5wt crystallized (0.2 M Li_3citrate, 20% PEG 3350) (Figure 83). The crystals were reproduced in custom 96-well grid-screens around this condition and optimized in 24-well hanging-drop setups to obtain larger crystals for data collection. The only difference compared to ATG5wt was, that for ATG5T75A, the crystallization setups had to be incubated at lower temperatures.

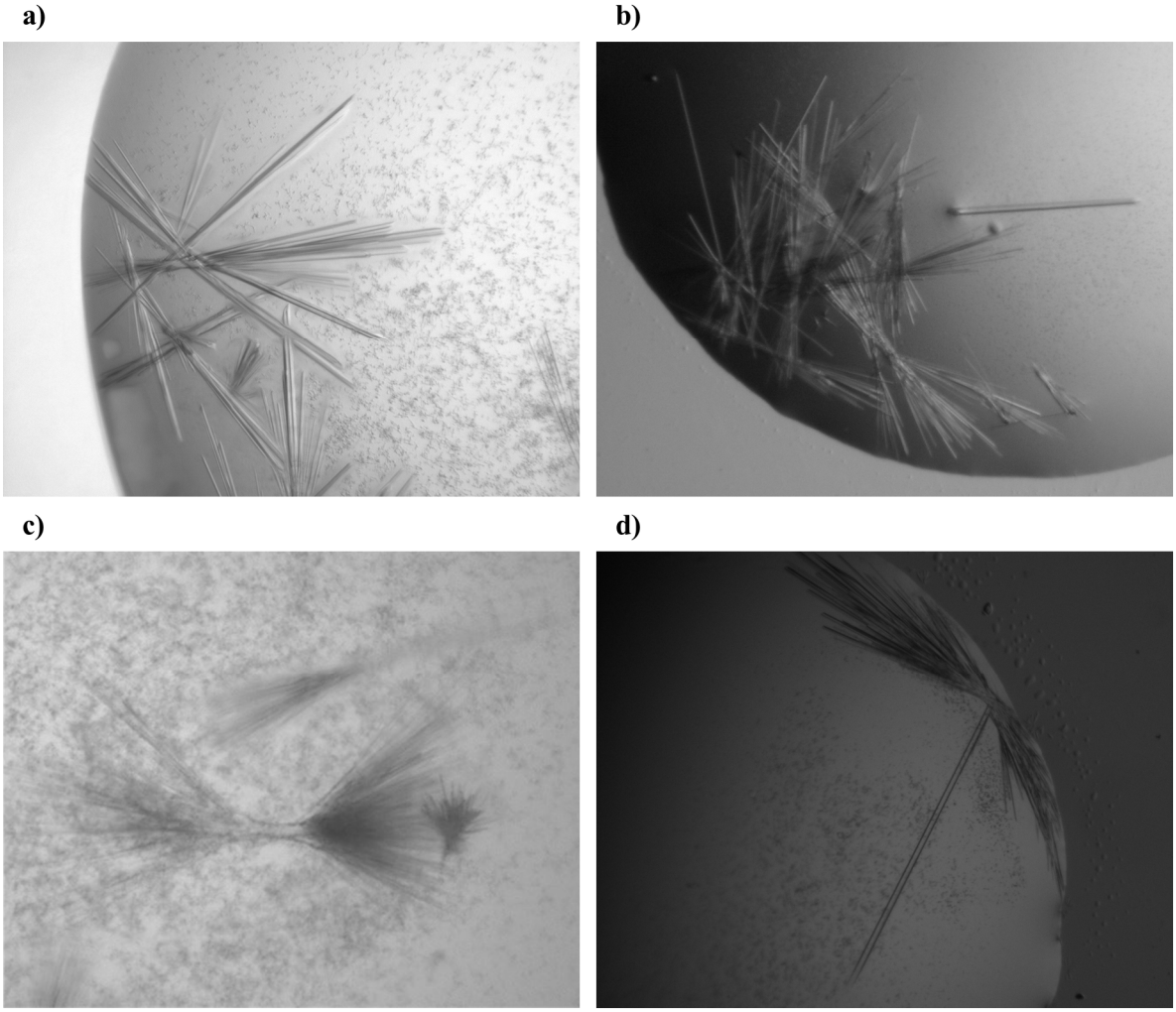


Figure 83: ATG5T75E and T75A crystals. a) T75E crystals reproduced in 96-well grid screen (0.35 M Li_3citrate, 35% PEG 3350, 5 mg/ml, 20°C), b) T75E crystals in hanging drop 24-well setups under optimized conditions (0.4 M Li_3citrate, 31% PEG 3350, 10 mg/ml, 20°C), c) T75A crystals reproduced in 96-well grid screen (0.35 M Li_3citrate, 35% PEG 3350, 5 mg/ml, 4°C), d) T75A crystals in hanging drop 24-well plate setups under optimized conditions (0.45 M Li_3citrate, 29% PEG 3350, 5 mg/ml, 12°C). The form of the crystals and the optimized conditions were very similar to the crystals of ATG5wt.

The crystals of both ATG5T75E and T75A diffracted to a resolution of around 4 Å in house. Data sets for both crystals were collected, using the Pilatus 6M detector of beamline P11 of the PETRA III synchrotron, at a wavelength of 1.033 and 1.000 Å respectively and a temperature of 100 K, where T75E diffracted to around 2.3 Å and T75A to around 2.7 Å. Like ATG5wt, both mutants crystallized in space group $P4_12_12$ with one molecule in the asymmetric unit. The structures were solved by molecular replacement, using the same model as for the wild type form (ATG5 of the human ATG5-12/16 complex from PDB-entry 4GDK). The structures of ATG5T75E and T75A are shown in Figure 84 and Figure 85.

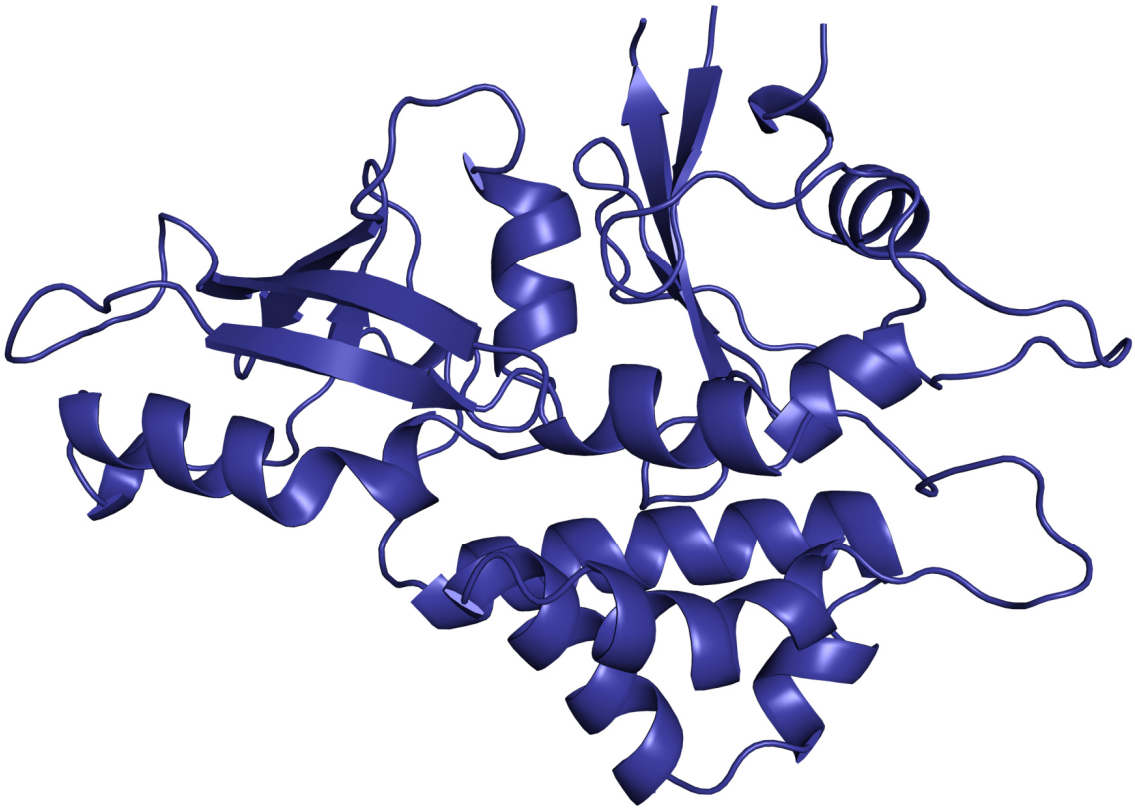


Figure 84: Crystal structure of ATG5T75E.

Figure 85: Crystal structure of ATG5T75A.

Table 20: Data collection and refinement statistics for ATG5T75E and T75A.

Data collection	**ATG5T75E**	**ATG5T75A**
X-ray source	P11, Petra III	P11, Petra III
Wavelength [Å]	1.033	1.000
Resolution [Å]	2.30 (2.38 – 2.30)	2.70 (2.80 – 2.70)
Space group	$P4_12_12$	$P4_12_12$
Unit-cell parameters [Å]	73.8, 73.8, 115.2 90.0, 90.0, 90.0	74.1, 74.1, 115.4 90.0, 90.0, 90.0
R_{merge}	0.119 (0.778)	0.132 (0.722)
R_{meas}	0.121 (0.794)	0.135 (0.736)
CC1/2	0.999 (0.953)	0.999 (0.954)
‹I/σ(I)›	23.70 (5.16)	26.37 (5.46)
Completeness [%]	99.9 (99.6)	100.0 (100.0)
No. of reflections	376765 (36314)	237959 (23696)
Multiplicity	25.6 (25.4)	25.5 (26.4)
Refinement		
Resolution [Å]	38.7 – 2.30	47.7 – 2.70
No. of reflections	14725 (1431)	9341 (898)
R_{work}	0.178 (0.233)	0.186 (0.263)
R_{free}	0.199 (0.285)	0.229 (0.359)
Asymmetric unit	1 molecule	1 molecule
No. of atoms		
Protein	2205	2195
Ligands	0	0
Water	81	46
Average *B* factor [Å^2]		
Protein	37.6	37.5
Ligands	/	/
Water	37.7	32.2
R.m.s. deviations		
Bond length [Å]	0.017	0.009
Bond angles [°]	1.68	1.39
Ramachandran favored [%]	99	99
Ramachandran outliers [%]	0.0	0.0

The two mutant variants of ATG5 were crystallized, but no differences in the structures compared to ATG5wt were identified. However it was still possible that the binding of ATG16L1 was influenced by the phosphorylation of T75, e.g. by a destabilization of the tertiary structure of ATG5. For this reason, thermofluor-assay assays and limited proteolysis experiments were performed, using ATG5wt, T75E and T75A.

4.3.3.2 ATG5 thermofluor-assay

To identify a possible influence of T75 phosphorylation on the stability of ATG5, a thermofluor-assay was performed, using ATG5wt, T75E and T75A (Figure 86). The melting temperatures of all ATG5 variants were in the range of 38-39°C. No significant influence of the phosphorylation of T75 on the thermostability of ATG5 was observed.

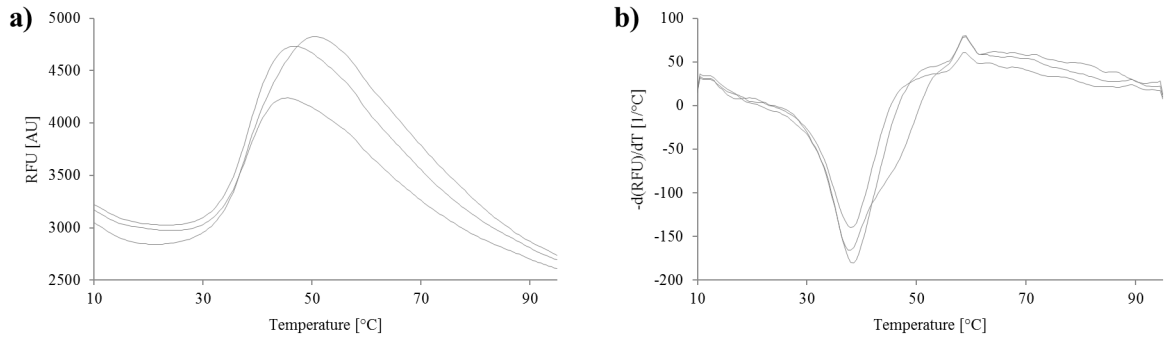


Figure 86: Thermofluor-assay for ATG5wt, T75E and T75A. a) Melting curves and b) melting peaks in 50 mM HEPES pH 7.0, 500 mM NaCl, 5% glycerol. All ATG5 variants had melting temperatures of 38-39°C.

4.3.3.3 ATG5 limited proteolysis

In a last experiment the stability of the different ATG5 variants was probed by proteolytic digestion experiments (Figure 87). If the tertiary structure of the mutants was destabilized, different proteolytic digestion patterns would have been expected, as parts of the protein would have been more flexible and accessible for the proteases. However, also in this experiment the different ATG5 variants behaved identical. No influence of the phosphorylation of T75 on the degradation patterns in the limited proteolysis experiment was observed.

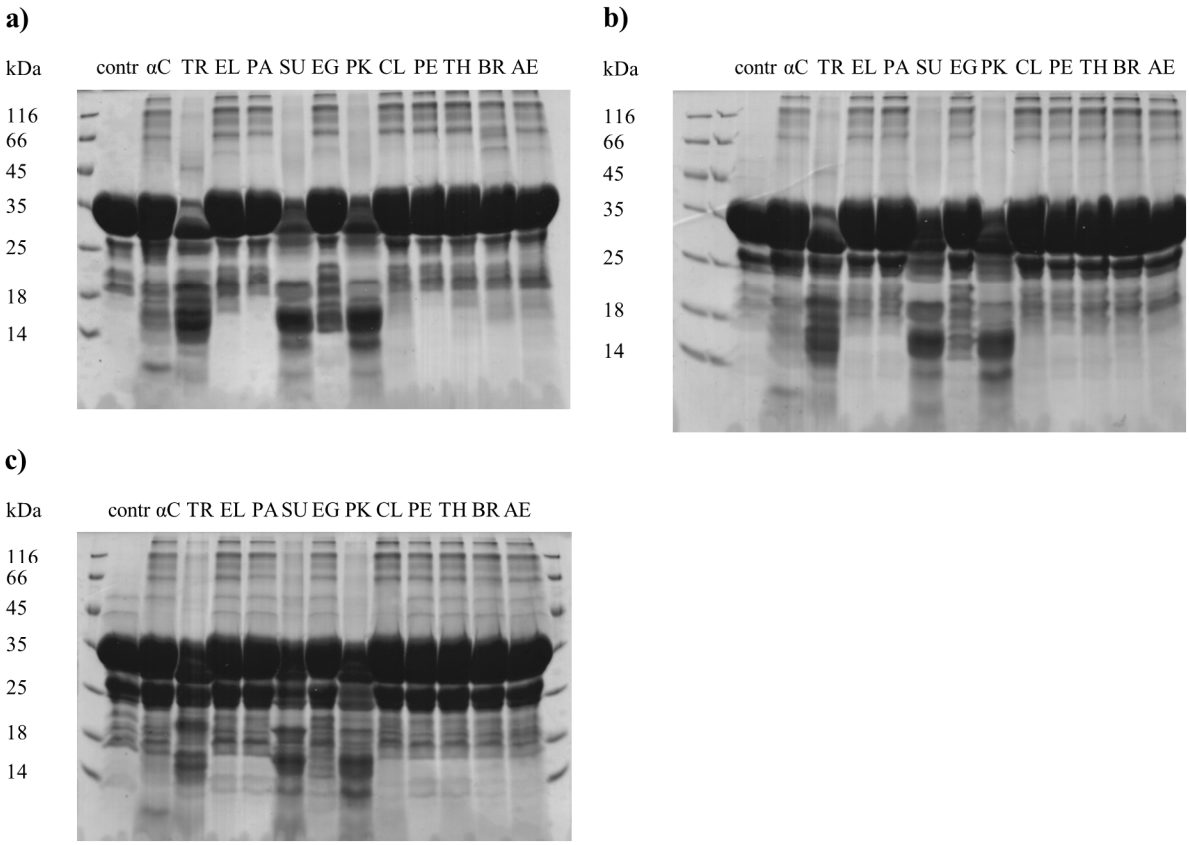


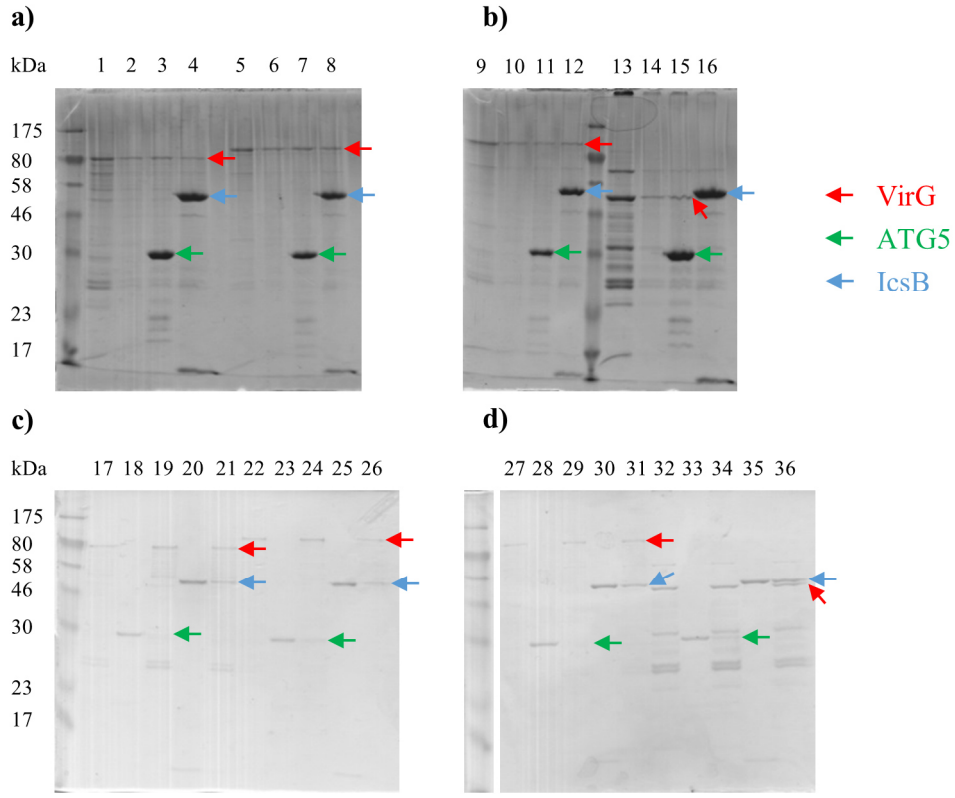
Figure 87: Coomassie stained SDS-gels for the limited proteolysis of a) ATG5wt, b) T75E and c) T75A, using following restriction enzymes from the Hampton Research Proti-Ace and Proti-Ace2 Kits: contr) no protease, αC) α-chymotrypsin, TR) trypsin, EL) elastase, PA) papain, SU) subtilisin, EG) endoproteinase Glu-C, PK) proteinase K, CL) clostripain, PE) pepsin, TH) thermolysin, BR) bromelain, AE) actinase. No differences in the degradation patterns were observed.

4.4 Interaction studies

4.4.1 VirG interactions

4.4.1.1 Interaction VirG – IcsB/ATG5

For a detailed understanding of the molecular basis of the *Shigella* camouflage mechanism, the previously published interactions of VirG with IcsB and ATG5 were tested in pulldown experiments, using GST-tagged VirGK2/K3/K4/K6 and His-tagged IcsBwt/ATG5wt (Figure 88). Both the GST- and the His-tag were used for pulldowns, but no interactions were observed. A major problem was the weak expression of the GST-VirG-constructs. At the same time weak protein bands were observed in the negative controls, making the interpretation of the results difficult, because of the impossibility to distinguish between weak protein bands, resulting from weak expression and from unspecific binding to the affinity beads.



	Beads	1. Protein	2. Protein
a)			
1	/	GST-VirGK2	/
2	His	GST-VirGK2	/
3	His	His-ATG5wt	GST-VirGK2
4	His	His-IcsBwt	GST-VirGK2
5	/	GST-VirGK3	/
6	His	GST-VirGK3	/
7	His	His-ATG5wt	GST-VirGK3
8	His	His-IcsBwt	GST-VirGK3

	Beads	1. Protein	2. Protein
b)			
9	/	GST-VirGK4	/
10	His	GST-VirGK4	/
11	His	His-ATG5wt	GST-VirGK4
12	His	His-IcsBwt	GST-VirGK4
13	/	GST-VirGK6	/
14	His	GST-VirGK6	/
15	His	His-ATG5wt	GST-VirGK6
16	His	His-IcsBwt	GST-VirGK6

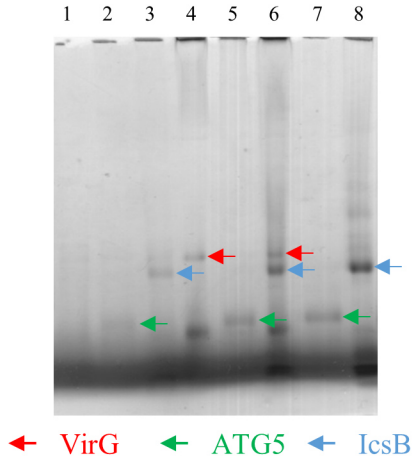
	Beads	1. Protein	2. Protein
c)			
17	GST	GST-VirGK2	/
18	GST	His-ATG5wt	/
19	GST	GST-VirGK2	His-ATG5wt
20	GST	His-IcsBwt	/
21	GST	GST-VirGK2	His-IcsBwt
22	GST	GST-VirGK3	/
23	GST	His-ATG5wt	/
24	GST	GST-VirGK3	His-ATG5wt
25	GST	His-IcsBwt	/
26	GST	GST-VirGK3	His-IcsBwt

	Beads	1. Protein	2. Protein
d)			
27	GST	GST-VirGK4	/
28	GST	His-ATG5wt	/
29	GST	GST-VirGK4	His-ATG5wt
30	GST	His-IcsBwt	/
31	GST	GST-VirGK4	His-IcsBwt
32	GST	GST-VirGK6	/
33	GST	His-ATG5wt	/
34	GST	GST-VirGK6	His-ATG5wt
35	GST	His-IcsBwt	/
36	GST	GST-VirGK6	His-IcsBwt

Figure 88: Coomassie stained SDS-gels for pulldown interaction studies, using GST-VirGK2/K3/K4/K6 with His-IcsBwt/ATG5wt. a) controls and His-pulldowns for VirGK2 and VirGK3, b) controls and His-pulldowns for VirGK4 and VirGK6, c) controls and GST-pulldowns for VirGK2 and VirGK3, d) controls and GST-pulldowns for VirGK4 and VirGK6. Red arrows mark the VirG-constructs, blue arrows IcsB and green arrows ATG5. No interactions were observed.

In another approach the interaction between GST-tagged VirGK4/K6 and His-tagged IcsBwt/ATG5wt was tested, using blue native PAGE (Figure 89). The samples were mixed and analyzed on a gel. If a

complex had been formed, the appearance of a higher MW band would have been expected. However, it was not possible to observe the interactions. Again one problem was the low amount of available VirG-protein, as VirGK4 could not be detected on the gel. All other proteins were observed as individual proteins, but no complexes were detected. The only sign of a possible interaction was the vanishing of the VirGK6 band, when ATG5 was added. On the other hand, if all VirG protein was in a complex with ATG5, the appearance of a new band with a considerably higher MW would have been expected.



	1. Protein	2. Protein
1	GST-VirGK4	/
2	GST-VirGK4	His-ATG5wt
3	GST-VirGK4	His-IcsBwt
4	GST-VirGK6	/
5	GST-VirGK6	His-ATG5wt
6	GST-VirGK6	His-IcsBwt
7	His-ATG5wt	/
8	His-IcsBwt	/

Figure 89: Blue native PAGE interaction studies, using GST-VirGK4/K6 with His-IcsBwt/ATG5wt. Red arrows mark the VirG-constructs, blue arrows mark IcsB and green arrows mark ATG5. No interactions were observed.

In the next step interaction of untagged VirGK3 and His-tagged IcsBwt/ATG5wt was tested, using an S200 16/60 column (Figure 90). The proteins were separated both individually and in all possible combinations. If a stable complex was formed, the emergence of an additional peak at a lower elution volume would have been expected. But again all proteins eluted in separate peaks and no additional peaks, corresponding to formed complexes, were observed. For individual VirGK3 two peaks were present. The higher peak with an elution volume of about 46 ml corresponded to VirGK3, the lower peak with an elution volume of about 81 ml corresponded to individual GST and TEV-protease that were separated in this run and were not present in later runs anymore.

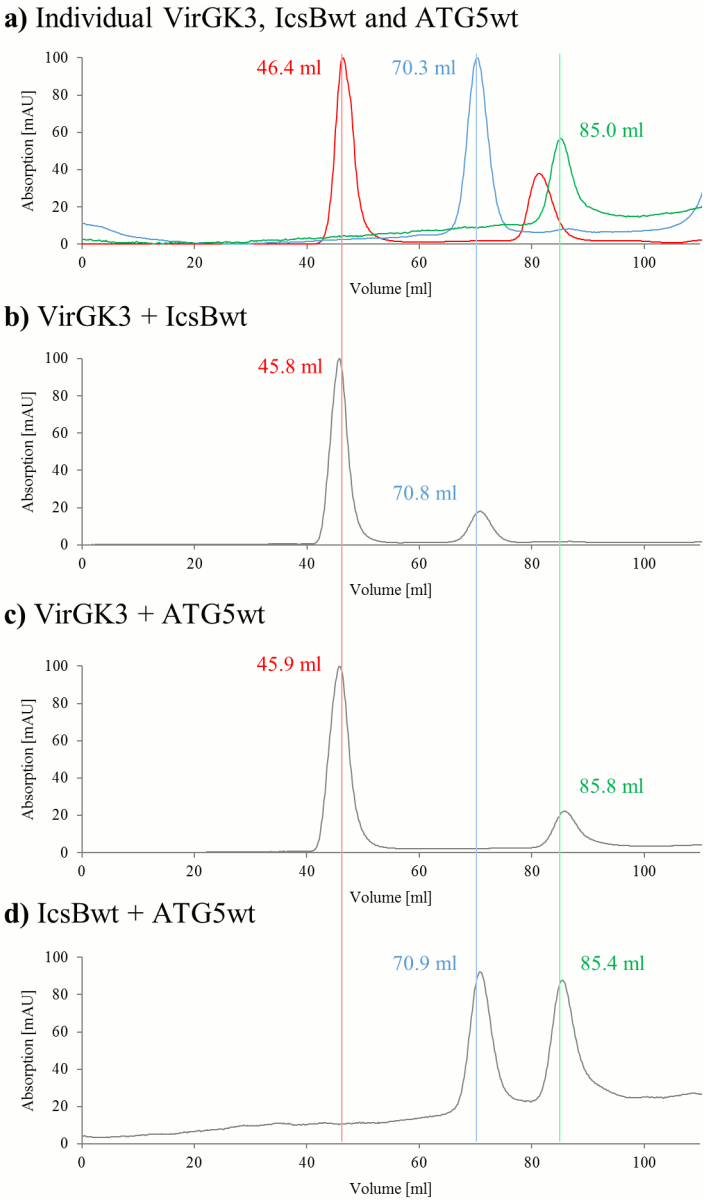
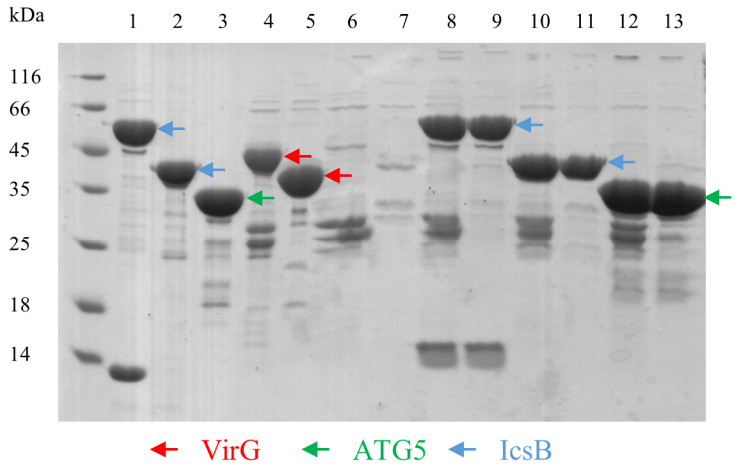


Figure 90: SEC elution profiles from an S200 16/60 column for a) individual VirGK3 (red), His-IcsBwt (blue) and His-ATG5wt (green); b) mixed VirGK3 and IcsBwt; c) mixed VirGK3 and ATG5wt; d) mixed IcsBwt and ATG5wt. Vertical colored lines mark the position of individual protein peaks. The elution volumes are indicated. For individual VirGK3 two peaks were present. The higher peak at around 46 ml corresponded to VirGK3, the lower peak at around 81 ml corresponded to individual GST and TEV-protease that were separated in this run and were not present in later runs. No interactions were observed.

In summary, using the constructs that were available before using the ESPRIT-system, the interactions of VirG to IcsB or ATG5 could not be observed. This might have been due to the formation of soluble aggregates by the VirG-constructs, as observed in SEC before, that obscured the interaction site. Another major problem was the low amount of available VirG protein. As the constructs that were obtained, using the ESPRIT-system, were expressed in higher quantities, soluble and presumably folded, further interaction studies were performed, using these new constructs.

The interactions between VirG and IcsB/ATG5 were tested again in pulldown experiments, using untagged VirGE94/E76 and His-tagged IcsBwt/IcsBK9/ATG5wt (Figure 91). As the VirG-constructs were untagged, only pulldowns, using the His-tags were done. All proteins were present in sufficient amounts and unspecific binding of the VirG-constructs to the His-affinity beads was insignificant. Still no interactions between these VirG-constructs and IcsB or ATG5 were observed.



	Beads	1. Protein	2. Protein
1	/	His-IcsBwt	/
2	/	His-IcsBK9	/
3	/	His-ATG5wt	/
4	/	VirGE76	/
5	/	VirGE94	/
6	His	VirGE76	/
7	His	VirGE94	/
8	His	His-IcsBwt	VirGE76
9	His	His-IcsBwt	VirGE94
10	His	His-IcsBK9	VirGE76
11	His	His-IcsBK9	VirGE94
12	His	His-ATG5wt	VirGE76
13	His	His-ATG5wt	VirGE94

Figure 91: Coomassie stained SDS-gel for pulldown interaction studies, using untagged VirGE94/E76 with His-tagged IcsBwt/IcsBK9/ATG5wt. Red arrows mark the VirG-constructs, blue arrows mark the IcsB-constructs and green arrows mark ATG5. No interactions were observed.

MST was used as a complementary method to verify these results. Labeled untagged VirGE94/E76 as well as unlabeled His-tagged IcsBwt/ATG5wt and GST as a negative control, were used in these experiments (Figure 92). For GST and ATG5 the curves showed a steep increase at high unlabeled protein concentrations. The reason for this might have been a higher viscosity of the solution, due to the very high protein concentration in the 0.1-1 mM range. No interactions of the VirG-constructs to GST or ATG5 were observed and also VirGE94 and IcsB did not interact. However, IcsB did interact with VirGE76 with a K_D of 600±205 nM.

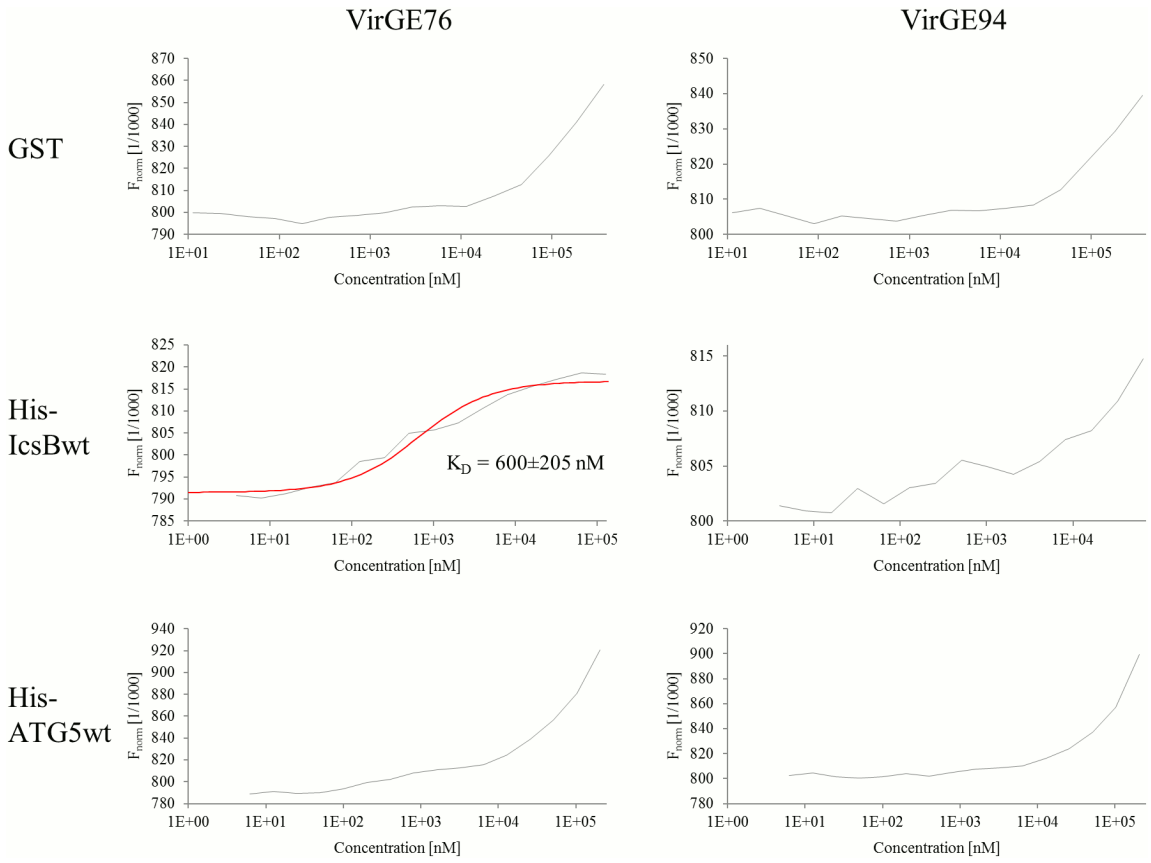
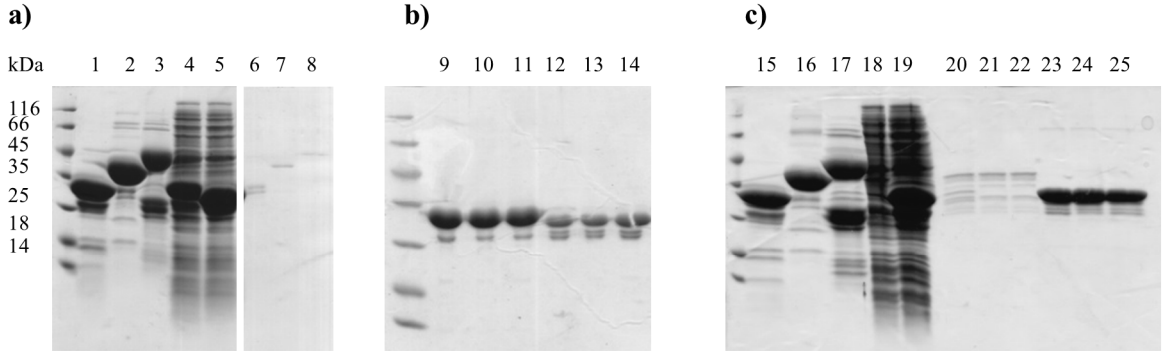


Figure 92: MST interaction studies, using labeled VirGE94/E76 with unlabeled His-IcsBwt/ATG5wt and GST as a negative control. Most curves showed a steep increase at high unlabeled protein concentrations. An interaction of IcsBwt with VirGE76 with a K_D of 600±205 nM, but no interaction of IcsBwt with VirGE94 or of ATG5 with one of the VirG-constructs, was observed.

4.4.1.2 Interaction VirG – N-WASP

N-WASP is a published interaction partner of VirG, which is essential for VirG's function in ABM. In several publications, binding of different domains of VirG to N-WASP has been shown [85], [135], [145], [194]. Using the soluble VirG-constructs, obtained from the ESPRIT system, the interaction of VirG to N-WASP should have been further characterized. The expression of full length N-WASP protein is problematic as it tends to aggregate [195]. Hence several individual N-WASP-domains and domain-combinations were expressed as GST-fusion proteins in *E. coli*. Pulldowns were performed, using purified untagged VirGE57/E94/E76 and cell lysates from *E. coli*, producing GST-tagged N-WASP-GBD/C/VVCA/VV (Figure 93). Expression of N-WASP-VVCA was weak, all other proteins were present in sufficient amounts. Unspecific interaction of the VirG-constructs with the GST-affinity beads was negligible. However, in the pulldowns only weak additional bands corresponding to the VirG-constructs were detected, which were similar in strength to the negative controls. Hence no interactions between the VirG- and N-WASP-constructs were observed in these pulldowns.



	Beads	1. Protein	2. Protein
a)			
1	/	VirGE57	/
2	/	VirGE94	/
3	/	VirGE76	/
4	/	GST-N-WASP-GBD	/
5	/	GST-N-WASP-C	/
6	GST	VirGE57	/
7	GST	VirGE94	/
8	GST	VirGE76	/

	Beads	1. Protein	2. Protein
b)			
9	GST	GST-N-WASP-GBD	VirGE57
10	GST	GST-N-WASP-GBD	VirGE94
11	GST	GST-N-WASP-GBD	VirGE76
12	GST	GST-N-WASP-C	VirGE57
13	GST	GST-N-WASP-C	VirGE94
14	GST	GST-N-WASP-C	VirGE76

	Beads	1. Protein	2. Protein
c)			
15	/	VirGE57	/
16	/	VirGE94	/
17	/	VirGE76	/
18	/	GST-N-WASP-VVCA	/
19	/	GST-N-WASP-VV	/
20	GST	GST-N-WASP-VVCA	VirGE57
21	GST	GST-N-WASP-VVCA	VirGE94
22	GST	GST-N-WASP-VVCA	VirGE76
23	GST	GST-N-WASP-VV	VirGE57
24	GST	GST-N-WASP-VV	VirGE94
25	GST	GST-N-WASP-VV	VirGE76

Figure 93: Coomassie stained SDS-gels for pulldown interaction studies, using VirGE57/E94/E76 with GST-N-WASP-GBD/C/VV/VVCA. a) controls, b) pulldowns with GST-N-WASP-GBD/C, c) controls and pulldowns with GST-N-WASP-VV/VVCA. No interactions were observed.

MST was used to confirm these results. Labeled VirGE94/E76 with unlabeled GST-N-WASP-GBD/C/VVCA/VV were used in these experiments (Figure 94). GST negative controls are shown in Figure 92. For none of the VirG-N-WASP combinations a clear sigmoidal binding curve was detected, hence no interactions were observed.

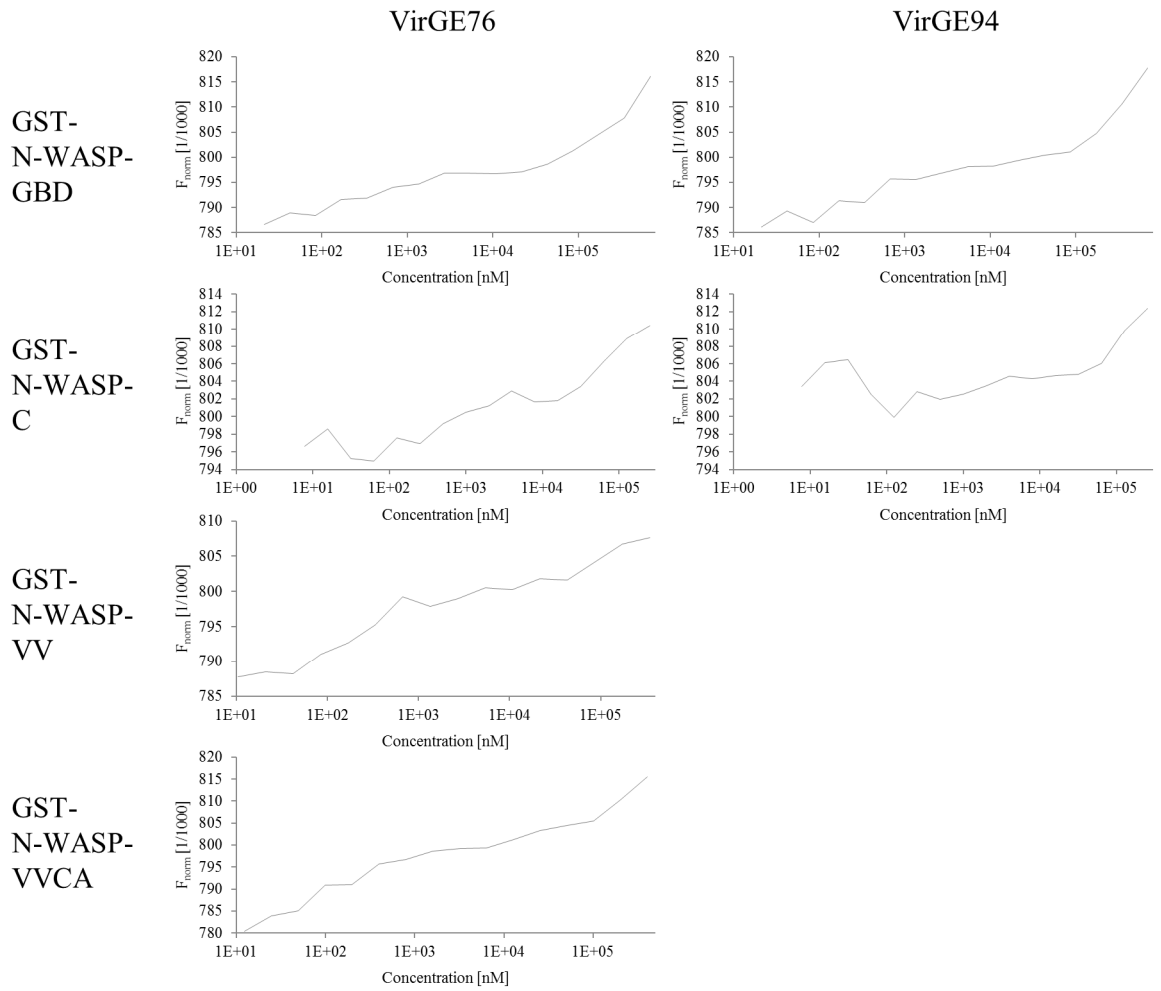
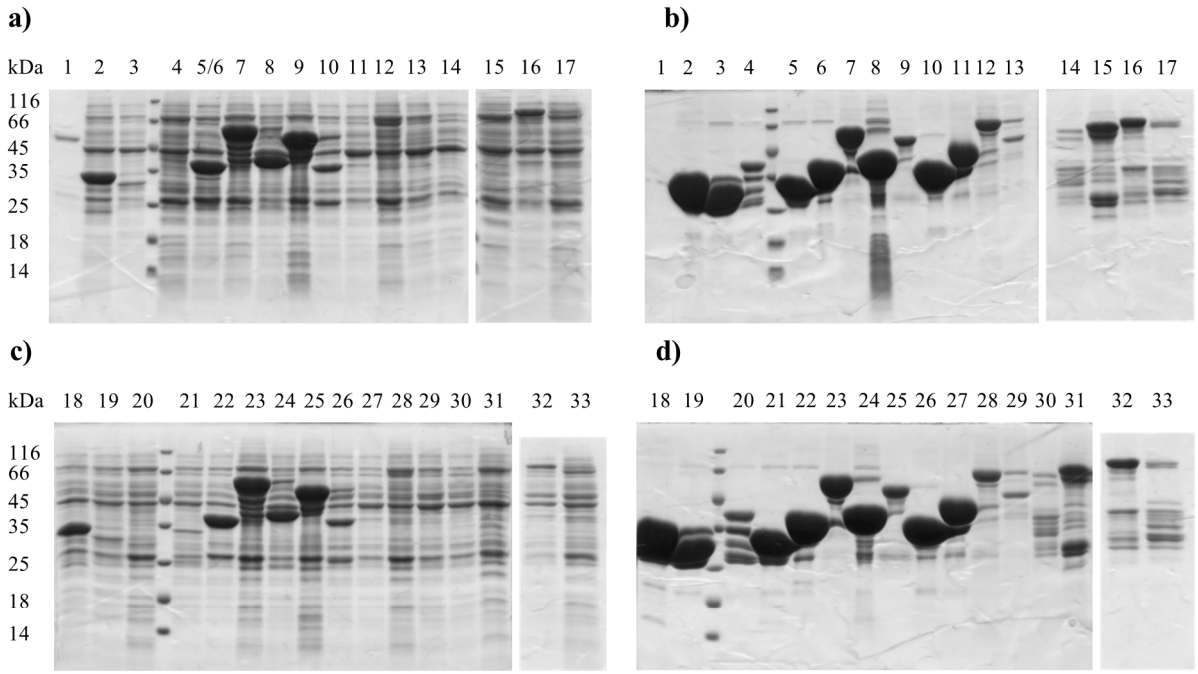


Figure 94: MST interaction studies, using labeled VirGE94/E76 with unlabeled GST-N-WASP-GBD/C/VV/VVCA. No sigmoidal binding curves were detected in these experiments, hence no interactions were observed.

4.4.1.3 Interaction VirG – other proteins

VirG is a protein that exploits the human actin polymerization system to permit ABM for *Shigella*. For this reason the interaction of VirGE76 with a number of proteins and domains of proteins, involved in or associated to the actin polymerization system, was tested. Pulldown studies were performed, using His-tagged VirGE76 with GST-tagged constructs of other proteins, comprising domains of N-WASP, Toca-1, Cip4, Cdc42 and a number of diaphanous formins (Figure 95). Some of the GST-tagged constructs, especially N-WASP-GBD and all Toca-1-constructs, bound unspecifically to His-affinity beads (Figure 95 a). Apart from this unspecific binding, no interactions between VirG and other protein constructs were observed.



	1. Protein
a) and b)	
1	His-VirGE76
2	GST-N-WASP-GBD
3	GST-N-WASP-C
4	GST-N-WASP-VVCA
5	GST-N-WASP-VV
6	GST-Toca-1K4
7	GST-Toca-1K5
8	GST-Toca-1K6
9	GST-Toca-1K7
10	GST-Cip4K1
11	GST-Cdc42
12	GST-Diaph2-GBD
13	GST-Diaph3-GBD
14	GST-Daam2-GBD
15	GST-Fhod1-GBD
16	GST-Fmnl2-GBD
17	GST-Fmnl3-GBD

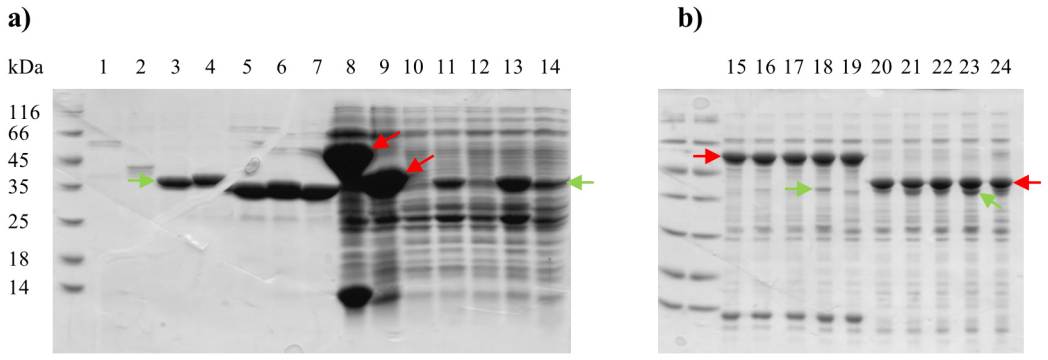
	1. Protein	2. Protein
c) and d)		
18	His-VirGE76	GST-N-WASP-GBD
19	His-VirGE76	GST-N-WASP-C
20	His-VirGE76	GST-N-WASP-VVCA
21	His-VirGE76	GST-N-WASP-VV
22	His-VirGE76	GST-Toca-1K4
23	His-VirGE76	GST-Toca-1K5
24	His-VirGE76	GST-Toca-1K6
25	His-VirGE76	GST-Toca-1K7
26	His-VirGE76	GST-Cip4K1
27	His-VirGE76	GST-Cdc42
28	His-VirGE76	GST-Diaph2-GBD
29	His-VirGE76	GST-Diaph3-GBD
30	His-VirGE76	GST-Daam2-GBD
31	His-VirGE76	GST-Fhod1-GBD
32	His-VirGE76	GST-Fmnl2-GBD
33	His-VirGE76	GST-Fmnl3-GBD

Figure 95: Coomassie stained SDS-gels for pulldown interaction studies, using His-tagged VirGE76 with different GST-tagged constructs of N-WASP, Toca-1, Cip4, Cdc42 and several diaphanous formins. a) controls His-pulldown, b) controls GST-pulldown, c) His-pulldowns, d) GST-pulldowns. No interactions were observed.

4.4.2 IcsB interactions

4.4.2.1 Interaction IcsB – Toca-1-HR1

IcsB has been shown to interact with the HR1 domain of Toca-1, which has a function in the activation of N-WASP. To verify the interaction of IcsB and Toca-1, a pulldown experiment was performed, using His-tagged IcsBwt (1-494, in complex with IpgA) and IcsBK9 (169-494, no IpgA), as well as four GST-tagged constructs of the Toca-1-HR1 domain (Toca-1K1-K4) and a GST-tagged construct of the HR1-domain of Cip4 (Cip4K1) (Figure 96). The unspecific binding of GST-Toca-1K2, K4 and GST-Cip4K1 to the His-affinity beads was problematic. Additionally the GST-constructs didn't bind properly to the GST-affinity beads, rendering this half of the experiment not interpretable. Another problem was that IcsBK9 and the Toca-1/Cip4-constructs had nearly the same MW, which makes the detection of an interaction between these two proteins difficult. In the pulldowns, using the His-tagged IcsB-constructs, only the unspecific interactions were detected that were already observed in the controls, hence no interaction between IcsB and Toca-1 was observed.



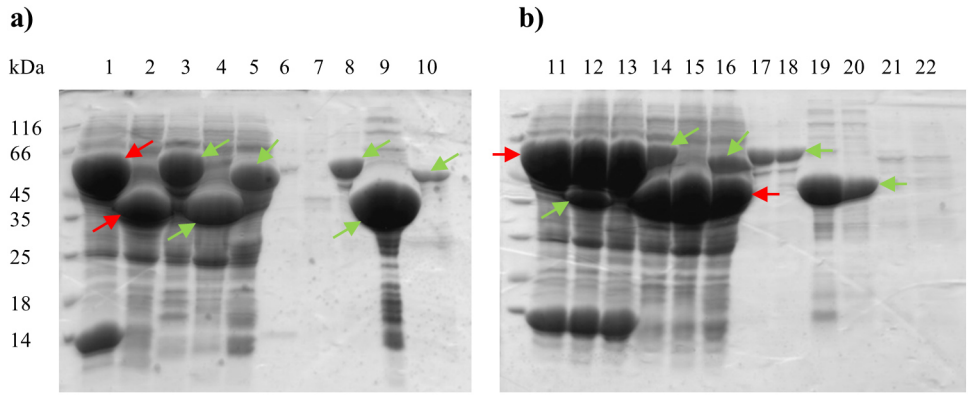
	Beads	1. Protein	2. Protein
a)			
1	GST	His-IcsBwt	/
2	GST	His-IcsBK9	/
3	GST	GST-Toca-1K1	/
4	GST	GST-Toca-1K2	/
5	GST	GST-Toca-1K3	/
6	GST	GST-Toca-1K4	/
7	GST	GST-Cip4K1	/
8	His	His-IcsBwt	/
9	His	His-IcsBK9	/
10	His	GST-Toca-1K1	/
11	His	GST-Toca-1K2	/
12	His	GST-Toca-1K3	/
13	His	GST-Toca-1K4	/
14	His	GST-Cip4K1	/

	Beads	1. Protein	2. Protein
b)			
15	His	His-IcsBwt	GST-Toca-1K1
16	His	His-IcsBwt	GST-Toca-1K2
17	His	His-IcsBwt	GST-Toca-1K3
18	His	His-IcsBwt	GST-Toca-1K4
19	His	His-IcsBwt	GST-Cip4K1
20	His	His-IcsBK9	GST-Toca-1K1
21	His	His-IcsBK9	GST-Toca-1K2
22	His	His-IcsBK9	GST-Toca-1K3
23	His	His-IcsBK9	GST-Toca-1K4
24	His	His-IcsBK9	GST-Cip4K1

	Beads	1. Protein	2. Protein
c)			
25	GST	GST-Toca-1K1	His-IcsBwt
26	GST	GST-Toca-1K1	His-IcsBK9
27	GST	GST-Toca-1K2	His-IcsBwt
28	GST	GST-Toca-1K2	His-IcsBK9
29	GST	GST-Toca-1K3	His-IcsBwt
30	GST	GST-Toca-1K3	His-IcsBK9
31	GST	GST-Toca-1K4	His-IcsBwt
32	GST	GST-Toca-1K4	His-IcsBK9
33	GST	GST-Cip4K1	His-IcsBwt
34	GST	GST-Cip4K1	His-IcsBK9

Figure 96: Coomassie stained SDS-gels for pulldown interaction studies, using His-tagged IcsBwt/K9 with four GST-tagged constructs of Toca-1 and one GST-tagged construct of Cip4. a) controls, b) His-pulldowns, c) GST-pulldowns. Red arrows mark the IcsB-constructs, green arrows mark the Toca-1- and Cip4-constructs. No interaction was observed.

As no clear interaction was detected, using the individual Toca-1-HR1 domain, in a second step pulldown experiments were performed, using His-tagged IcsBwt and IcsBK9, as well as the Toca-1-constructs that were originally used to show this interaction (Toca-1K5-K7) [159] (Figure 97). All three GST-tagged Toca-1-constructs bound strongly and unspecifically to His-affinity beads, rendering the His-pulldowns not interpretable. In the GST-pulldowns, no additional bands corresponding to the IcsB-constructs were detected, hence no interactions between IcsB and these Toca-1-constructs were observed.



	Beads	1. Protein	2. Protein
a)			
1	His	His-IcsBwt	/
2	His	His-IcsBK9	/
3	His	GST-Toca-1K5	/
4	His	GST-Toca-1K6	/
5	His	GST-Toca-1K7	/
6	GST	His-IcsBwt	/
7	GST	His-IcsBK9	/
8	GST	GST-Toca-1K5	/
9	GST	GST-Toca-1K6	/
10	GST	GST-Toca-1K7	/

	Beads	1. Protein	2. Protein
b)			
11	His	His-IcsBwt	GST-Toca-1K5
12	His	His-IcsBwt	GST-Toca-1K6
13	His	His-IcsBwt	GST-Toca-1K7
14	His	His-IcsBK9	GST-Toca-1K5
15	His	His-IcsBK9	GST-Toca-1K6
16	His	His-IcsBK9	GST-Toca-1K7
17	GST	GST-Toca-1K5	His-IcsBwt
18	GST	GST-Toca-1K5	His-IcsBK9
19	GST	GST-Toca-1K6	His-IcsBwt
20	GST	GST-Toca-1K6	His-IcsBK9
21	GST	GST-Toca-1K7	His-IcsBwt
22	GST	GST-Toca-1K7	His-IcsBK9

Figure 97: Coomassie stained SDS-gels for pulldown interaction studies, using His-tagged IcsBwt/K9 with different GST-tagged constructs of Toca-1. a) controls, b) His- and GST-pulldowns. Red arrows mark the IcsB-constructs, green arrows mark the Toca-1-constructs. No interactions were observed.

Due to the problems in these experiments, the interaction of IcsBwt with a reduced number of Toca-1-constructs, was reassessed in a later experiment, together with other proteins (Figure 103). Also in this pulldown experiment, the used GST-tagged Toca-1-constructs showed a strong unspecific binding to His-affinity beads and no interactions were observed.

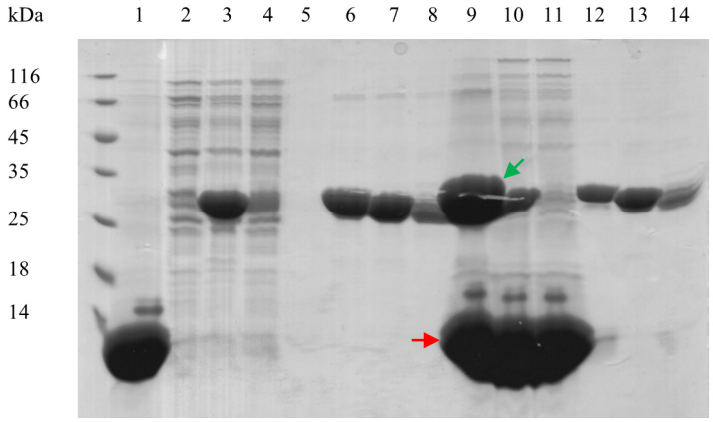
4.4.2.2 Interaction IcsB – Cholesterol

IcsB had been shown to interact with cholesterol through its cholesterol binding domain, located between residues 288-351 [154]. This interaction was tested, using dehydroergosterol (DHE), a fluorescent cholesterol analogon. Cholesterol and DHE are both virtually insoluble in water, hence DHE was solubilized, using methyl-β-cyclodextrin, a cyclic compound composed of seven glucose moieties, which can harbor one molecule of cholesterol [196]. The solubilized DHE was mixed with IcsBwt (in complex with IpgA), incubated and separated on an S200 10/30 column. It was not possible to detect the absorption of DHE at 324 nm. Possible reasons for this result might be that the solubilization of DHE did not work or that DHE did not bind to IcsB under the used conditions. In the former case DHE might not have been available for binding to IcsB, in the latter its fluorescence would have been quenched, due to the aqueous environment. Consequently the interaction of IcsB with DHE could not be verified.

4.4.2.3 Interaction IcsB – N-WASP

Phyre2 [197] was used to predict the tertiary structure of IcsB. The quality of the resulting model was poor, as no homologues structures were known. Still, a stretch in IcsB, comprising residues 99-130, was identified to show homology to a stretch in *E. coli* $EspF_U$, comprising residues 73-102, where it was responsible for the activation of N-WASP, by binding to the N-WASP-GBD [198]. To test if in IcsB this stretch also interacted with the N-WASP-GBD, two IcsB-constructs were prepared, the first covering the IcsB region corresponding to the $EspF_U$ N-WASP binding region (IcsBK30, 104-130), the

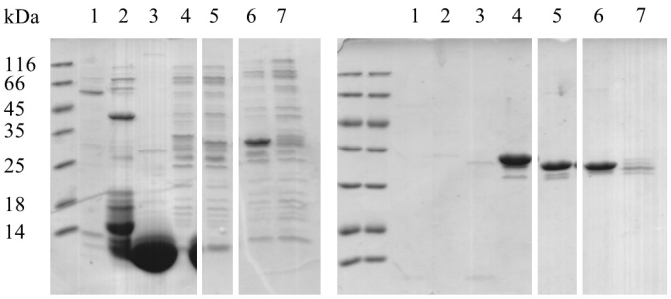
second covering additional regions N- and C-terminal of the binding region (IcsBK31, 99-136). Pulldown experiments were performed, using His-tagged N-WASP-GBD, as well as GST-tagged IcsBK30, IcsBK31 and N-WASP-C as a positive control (Figure 98). GST-IcsBK30, but not GST-IcsBK31, bound unspecifically to the His-affinity beads. The GBD and the C-domain of N-WASP, which served as a positive control for binding, clearly interacted in the His-pulldowns, while in the GST-pulldowns no interaction was observed. This might indicate a sterical problem for the interaction of the GBD and the C-domain, upon binding of the GST-tag to the GST-affinity beads. Between the tested IcsB-constructs and the N-WASP-GBD, no interactions were observed.



	Beads	1. Protein	2. Protein
1	His	His-N-WASP-GBD	/
2	His	GST-N-WASP-C	/
3	His	GST-IcsBK30	/
4	His	GST-IcsBK31	/
5	GST	His-N-WASP-GBD	/
6	GST	GST-N-WASP-C	/
7	GST	GST-IcsBK30	/
8	GST	GST-IcsBK31	/
9	His	His-N-WASP-GBD	GST-N-WASP-C
10	His	His-N-WASP-GBD	GST-IcsBK30
11	His	His-N-WASP-GBD	GST-IcsBK31
12	GST	GST-N-WASP-C	His-N-WASP-GBD
13	GST	GST-IcsBK30	His-N-WASP-GBD
14	GST	GST-IcsBK31	His-N-WASP-GBD

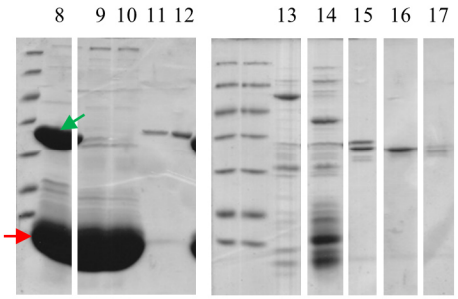
Figure 98: Coomassie stained SDS-gel for pulldown interaction studies, using His-tagged N-WASP-GBD with GST-tagged IcsBK30/K31/N-WASP-C. Except for the binding of N-WASP-GBD (red arrow) to N-WASP-C (green arrow), no interactions were observed.

a)



	1. Protein
a)	
1	His-IcsBwt
2	His-IcsBK9
3	His-N-WASP-GBD
4	GST-N-WASP-GBD
5	GST-N-WASP-C
6	GST-IcsBK30
7	GST-IcsBK31

b)



	Beads	1. Protein	2. Protein
b)			
8	His	His-N-WASP-GBD	GST-N-WASP-C
9	His	His-N-WASP-GBD	GST-IcsBK30
10	His	His-N-WASP-GBD	GST-IcsBK31
11	GST	GST-N-WASP-GBD	His-IcsBwt
12	GST	GST-N-WASP-GBD	His-IcsBK9
13	His	His-IcsBwt	GST-N-WASP-GBD
14	His	His-IcsBK9	GST-N-WASP-GBD
15	GST	GST-N-WASP-C	His-N-WASP-GBD
16	GST	GST-IcsBK30	His-N-WASP-GBD
17	GST	GST-IcsBK31	His-N-WASP-GBD

Figure 99: Coomassie stained SDS-gels for pulldown interaction studies, using GST- or His-tagged N-WASP-GBD with His-tagged IcsBwt/IcsBK9 or GST-tagged IcsBK30/K31/N-WASP-C. a) controls, one gel with His-, one with GST-pulldowns. b) pulldowns. Except for the binding of N-WASP-GBD (red arrow) to N-WASP-C (green arrow), no interactions were observed.

One problem might have been that the IcsB-constructs were too limited and other regions were important for the interaction. For this reason in a follow-up experiment His-tagged IcsBwt (1-494, in complex with IpgA) and IcsBK9 (169-494, no IcsB) were tested for binding to the N-WASP-GBD (Figure 99). GST-IcsBK30 again bound unspecifically to the His-affinity beads. Another problem was the weak expression of IcsBwt, IcsBK9 and IcsBK31. Binding was detected between the N-WASP-GBD and C-domain, which served as a positive control again. However, no interactions were observed between the N-WASP-GBD and the IcsB-constructs.

In a last experiment the interaction of IcsB and the N-WASP-GBD was tested, using a complementary method. Normally the N-WASP-GBD is disordered and only folds upon binding of an interaction partner, like the N-WASP-C domain or $EspF_U$. This property was used in a thermofluor-assay. For an unfolded N-WASP-GBD, no distinct melting curve or a low melting temperature would be expected, while for the folded domain a distinct melting curve with a higher melting point would be expected. Three different fusion proteins of N-WASP-GBD and IcsBK30/IcsBK31/N-WASP-C were created, based on the constructs published in [198]. The GBD-C fusion protein served as a positive control for binding, while the individual GBD served as a negative control. First the constructs were purified. Unfortunately the GBD-C fusion protein, the positive control for binding, was expressed poorly and could not be purified in sufficient amounts, even though it showed reasonable expression levels in test-expressions (Figure 100).

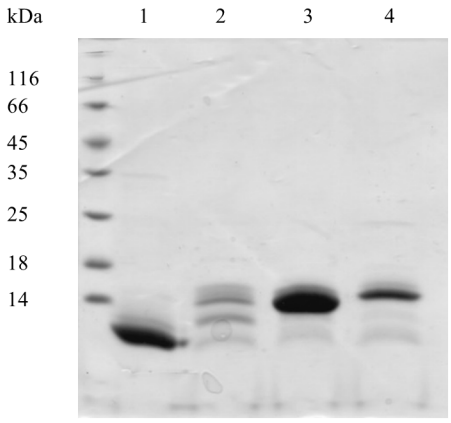
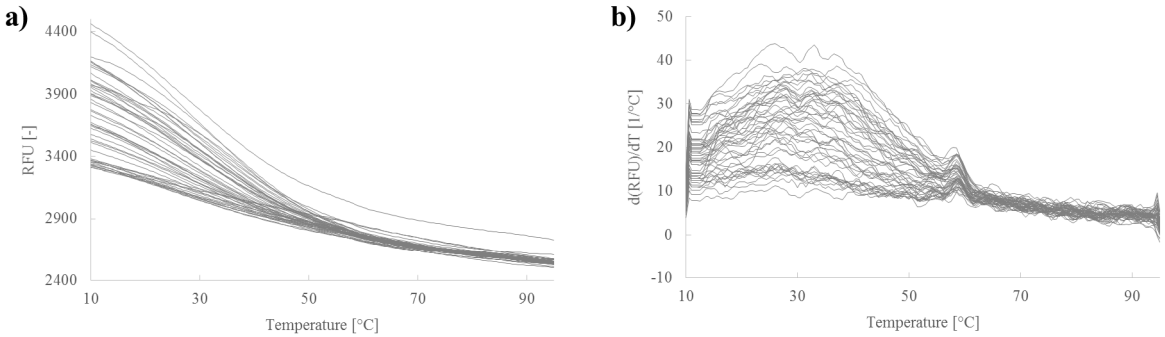


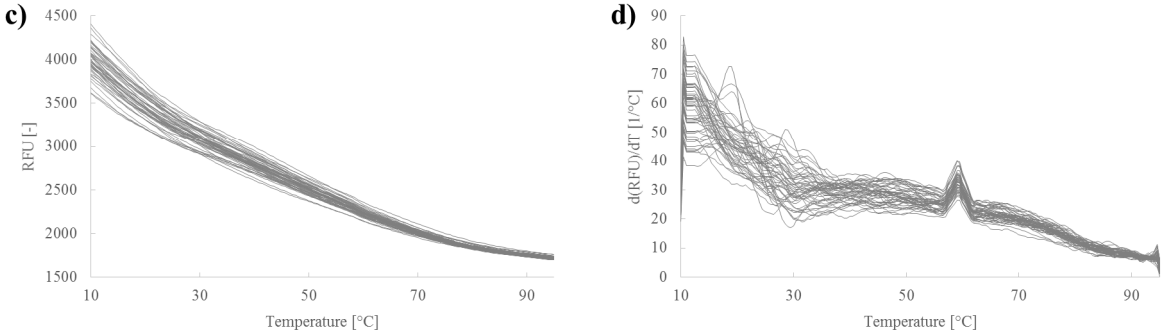
Figure 100: Coomassie stained SDS-gel of different N-WASP-GBD fusion proteins after His-affinity purification. 1) N-WASP-GBD, 2) N-WASP-GBD–N-WASP-C, 3) N-WASP-GBD–IcsBK30, 4) N-WASP-GBD–IcsBK31.

The purified proteins were used in a thermofluor-assay (Figure 101). The individual N-WASP-GBD and the N-WASP-GBD–IcsBK30 fusion protein didn't show distinct melting curves, hence their GBDs were probably unfolded. But the melting curve of the N-WASP-GBD–IcsBK31 fusion protein showed a slight inflection point, corresponding to a melting temperature of about 50°C. This indicated that the longer IcsBK31 construct might have interacted with the GBD of N-WASP and stabilized it. However, without the positive control for binding, it was difficult to evaluate the plausibility of the results.

N-WASP-GBD:



N-WASP-GBD–IcsBK30:



N-WASP-GBD–IcsBK31:

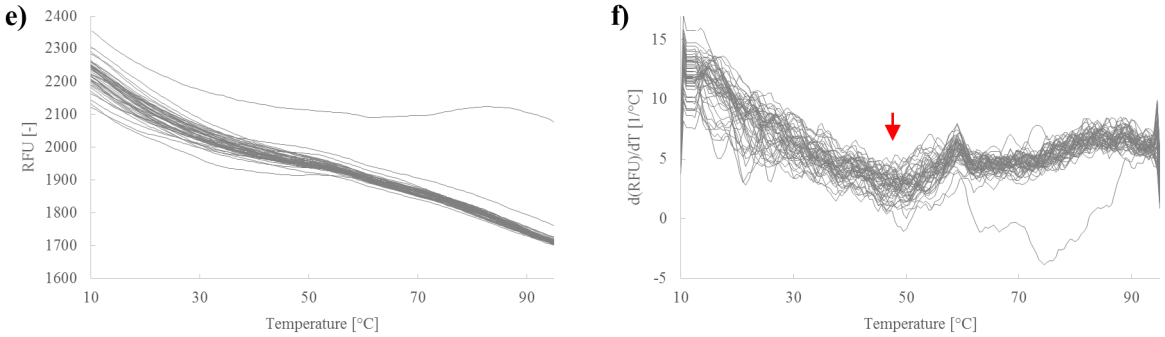
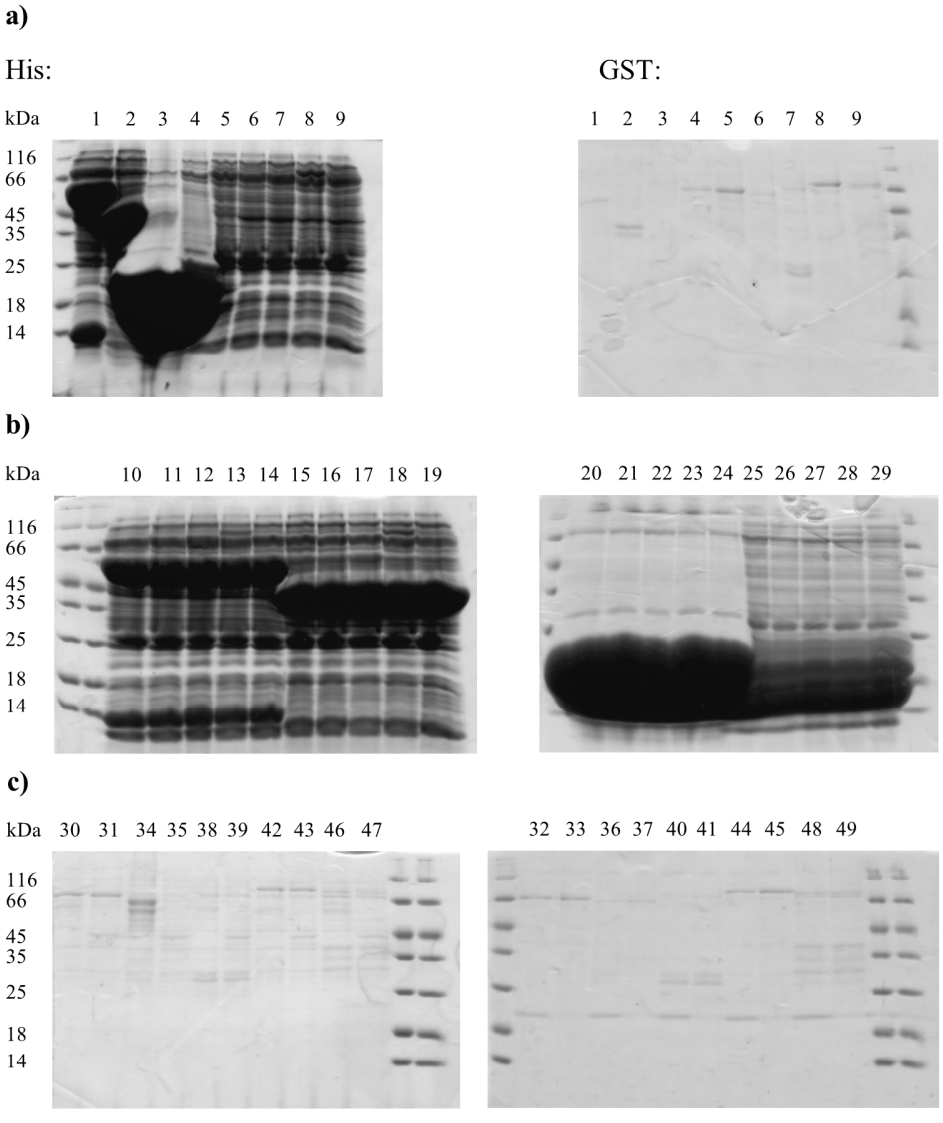


Figure 101: Thermofluor-assay with different N-WASP-GBD fusion proteins. For all proteins a buffer screen was performed. a) Melting curves and b) melting peaks for N-WASP-GBD. c) Melting curves and d) melting peaks for N-WASP-GBD–IcsBK30 fusion protein. e) Melting curves and f) melting peaks for N-WASP-GBD–IcsBK31 fusion protein. GBD–IcsBK31 showed a slight inflection point (red arrow), corresponding to a melting temperature of about 50°C.

4.4.2.4 Interaction IcsB – Diaphanous formins

Diaphanous formins contain a GBD and are autoinhibited in a similar way as N-WASP. Diaph1 and Diaph2 are reported to be essential for protrusion formation and IpgB2 has been shown to activate them [90]. As IcsB might have interacted with the N-WASP-GBD, here the interaction of His-/ZZ-His-tagged IcsBwt, IcsBK9, K30 and K31 with the GST-tagged GBDs of five different diaphanous formins were tested in pulldown-experiments (Figure 102). Unfortunately the GBD of Diaph1, which has been shown to be involved in the formation of *Shigella* protrusions [90], was not included in this experiment, due to prolonged cloning problems. Generally, the expression of the GBDs was weak. GST-Fmnl2-GBD was produced slightly better than the other constructs, but bound unspecifically to the His-affinity beads. No interactions between the IcsB-construct and the diaphanous formin GBDs were observed.



	1. Protein
a)	
1	His-IcsBwt
2	His-IcsBK9
3	ZZHis-IcsBK30
4	ZZHis-IcsBK31
5	GST-Diaph2-GBD
6	GST-Daam2-GBD
7	GST-Fhod1-GBD
8	GST-Fmnl2-GBD
9	GST-Fmnl3-GBD

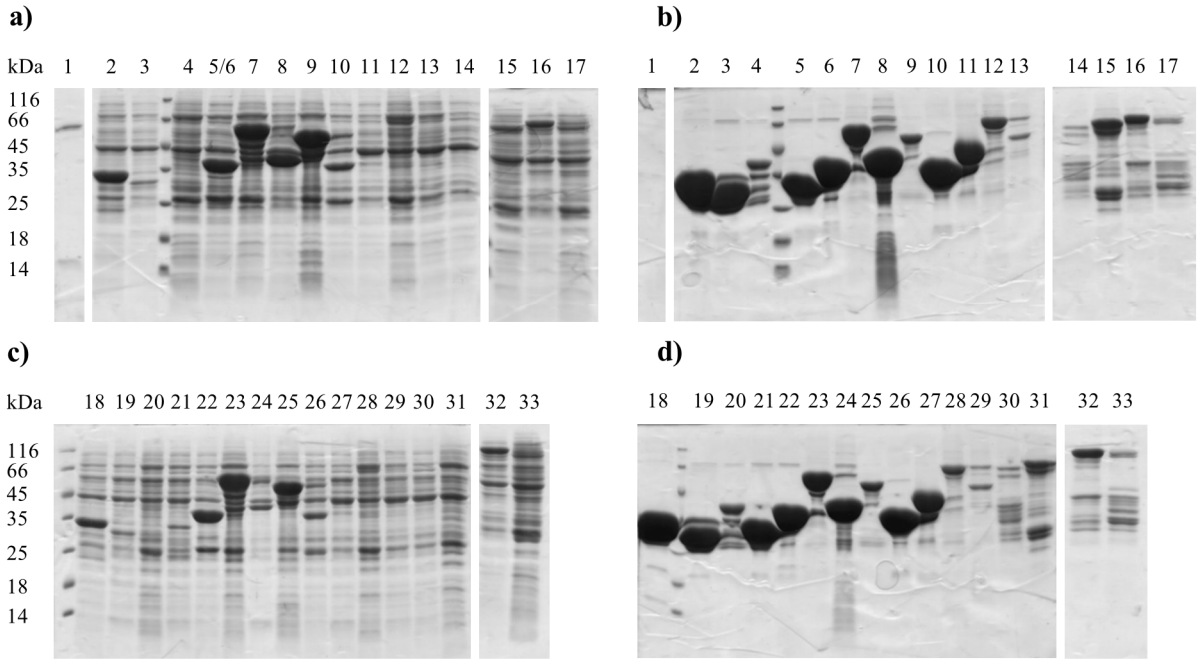
	Beads	1. Protein	2. Protein
b)			
10	His	His-IcsBwt	GST-Diaph2-GBD
11	His	His-IcsBwt	GST-Daam2-GBD
12	His	His-IcsBwt	GST-Fhod1-GBD
13	His	His-IcsBwt	GST-Fmnl2-GBD
14	His	His-IcsBwt	GST-Fmnl3-GBD
15	His	His-IcsBK9	GST-Diaph2-GBD
16	His	His-IcsBK9	GST-Daam2-GBD
17	His	His-IcsBK9	GST-Fhod1-GBD
18	His	His-IcsBK9	GST-Fmnl2-GBD
19	His	His-IcsBK9	GST-Fmnl3-GBD
20	His	ZZHis-IcsBK30	GST-Diaph2-GBD
21	His	ZZHis-IcsBK30	GST-Daam2-GBD
22	His	ZZHis-IcsBK30	GST-Fhod1-GBD
23	His	ZZHis-IcsBK30	GST-Fmnl2-GBD
24	His	ZZHis-IcsBK30	GST-Fmnl3-GBD
25	His	ZZHis-IcsBK31	GST-Diaph2-GBD
26	His	ZZHis-IcsBK31	GST-Daam2-GBD
27	His	ZZHis-IcsBK31	GST-Fhod1-GBD
28	His	ZZHis-IcsBK31	GST-Fmnl2-GBD
29	His	ZZHis-IcsBK31	GST-Fmnl3-GBD

	Beads	1. Protein	2. Protein
c)			
30	GST	GST-Diaph2-GBD	His-IcsBwt
31	GST	GST-Diaph2-GBD	His-IcsBK9
32	GST	GST-Diaph2-GBD	ZZHis-IcsBK30
33	GST	GST-Diaph2-GBD	ZZHis-IcsBK31
34	GST	GST-Daam2-GBD	His-IcsBwt
35	GST	GST-Daam2-GBD	His-IcsBK9
36	GST	GST-Daam2-GBD	ZZHis-IcsBK30
37	GST	GST-Daam2-GBD	ZZHis-IcsBK31
38	GST	GST-Fhod1-GBD	His-IcsBwt
39	GST	GST-Fhod1-GBD	His-IcsBK9
40	GST	GST-Fhod1-GBD	ZZHis-IcsBK30
41	GST	GST-Fhod1-GBD	ZZHis-IcsBK31
42	GST	GST-Fmnl2-GBD	His-IcsBwt
43	GST	GST-Fmnl2-GBD	His-IcsBK9
44	GST	GST-Fmnl2-GBD	ZZHis-IcsBK30
45	GST	GST-Fmnl2-GBD	ZZHis-IcsBK31
46	GST	GST-Fmnl3-GBD	His-IcsBwt
47	GST	GST-Fmnl3-GBD	His-IcsBK9
48	GST	GST-Fmnl3-GBD	ZZHis-IcsBK30
49	GST	GST-Fmnl3-GBD	ZZHis-IcsBK31

Figure 102: Coomassie stained SDS-gels for pulldown interaction studies, using His-tagged IcsBwt/IcsBK9 and ZZHis-tagged IcsBK30, K31 with GST-tagged constructs of Diaph2/Daam2/Fhod1/Fmnl2/Fmnl3. a) controls, one gel with His-, one with GST-pulldowns, b) His-pulldowns, c) GST-pulldowns. Except for the unspecific binding of GST-Fmnl2-GBD to the His-affinity beads, no interactions were observed.

4.4.2.5 Interaction IcsB – other proteins

In a last pulldown experiment, possible interactions of His-tagged IcsBwt with GST-tagged Diaph3-GBD, a diaphanous formin, and Cdc42, a small Rho-GTPase that controls the turnover of the actin-cytoskeleton, were tested. At the same time, some of the already tested interactions were reassessed, to confirm the results of the previous pulldown experiments. However, no interactions between IcsBwt and any of the tested constructs were observed (Figure 103).



	1. Protein
a), b)	
1	His-IcsBwt
2	GST-N-WASP-GBD
3	GST-N-WASP-C
4	GST-N-WASP-VVCA
5	GST-N-WASP-VV
6	GST-Toca-1K4
7	GST-Toca-1K5
8	GST-Toca-1K6
9	GST-Toca-1K7
10	GST-Cip4K1
11	GST-Cdc42
12	GST-Diaph2-GBD
13	GST-Diaph3-GBD
14	GST-Daam2-GBD
15	GST-Fhod1-GBD
16	GST-Fmnl2-GBD
17	GST-Fmnl3-GBD

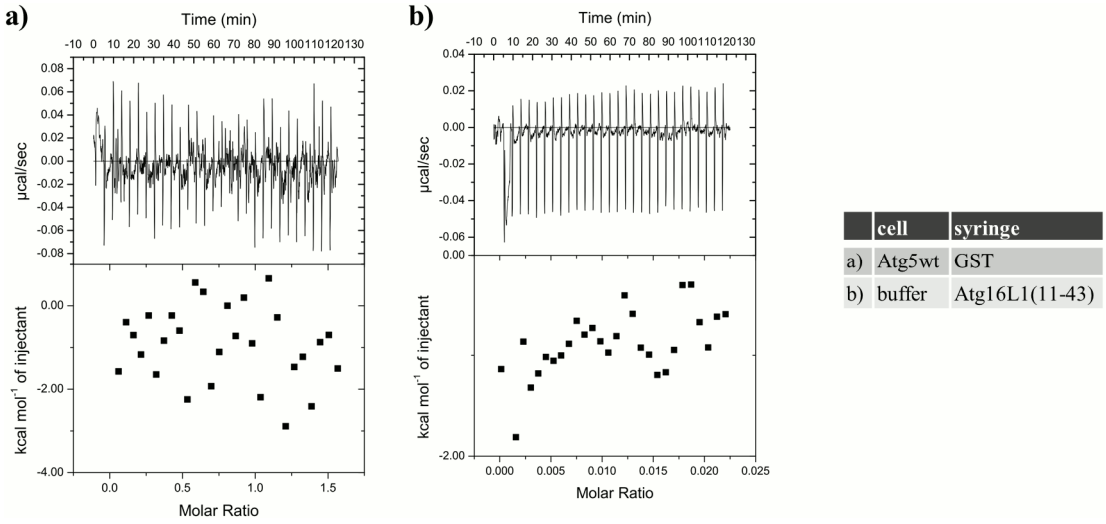
	1. Protein	2. Protein
c), d)		
18	His-IcsBwt	GST-N-WASP-GBD
19	His-IcsBwt	GST-N-WASP-C
20	His-IcsBwt	GST-N-WASP-VVCA
21	His-IcsBwt	GST-N-WASP-VV
22	His-IcsBwt	GST-Toca-1K4
23	His-IcsBwt	GST-Toca-1K5
24	His-IcsBwt	GST-Toca-1K6
25	His-IcsBwt	GST-Toca-1K7
26	His-IcsBwt	GST-Cip4K1
27	His-IcsBwt	GST-Cdc42
28	His-IcsBwt	GST-Diaph2-GBD
29	His-IcsBwt	GST-Diaph3-GBD
30	His-IcsBwt	GST-Daam2-GBD
31	His-IcsBwt	GST-Fhod1-GBD
32	His-IcsBwt	GST-Fmnl2-GBD
33	His-IcsBwt	GST-Fmnl3-GBD

Figure 103: Coomassie stained SDS-gels for pulldown interaction studies, using His-IcsBwt with different GST-tagged constructs of N-WASP/Toca-1/Cip4/Cdc42 and several diaphanous formins. a) controls, His-pulldown, b) controls, GST-pulldown, c) His-pulldowns, d) GST-pulldowns. No interactions were observed.

4.4.3 Interaction ATG5 – ATG16L1

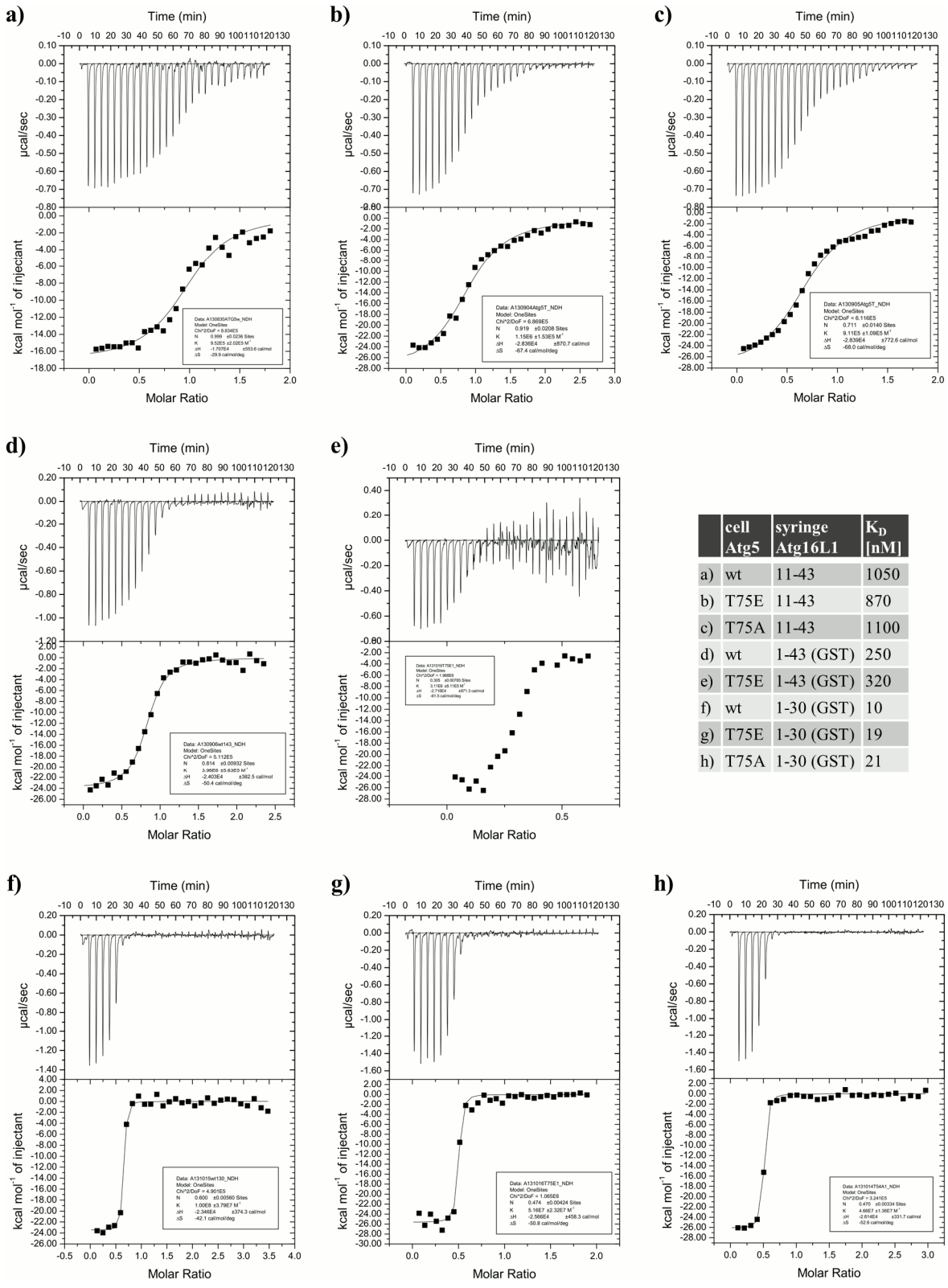
It was hypothized that the phosphorylation of ATG5 residue T75 might have an influence on the affinity of ATG5 to ATG16L1 or TECPR-1. For this reason, here the interaction of ATG5 and ATG16L1 was tested, using ITC and BLI. ATG5wt, T75E and T75A, as well as a synthetic ATG16L1 peptide, comprising residues 11-43, and two GST-ATG16L1 fusion proteins, comprising residues 1-30 and 1-43, were used in these experiments.

The results of the ITC experiments are shown in Figure 105. Clearly, in all experiments interaction of ATG5 and ATG16L1 was shown. The ATG16L1(11-43) peptide interacted with all forms of ATG5 with an affinity of about 1 µM (Figure 105 a-c). For GST-ATG16L1(1-43) the interaction with ATG5T75A was not assessed. Binding affinities to ATG5wt and T75E were in the range of 200-300 nM (Figure 105 d-e). Finally GST-ATG16L1(1-30) interacted with all variants of ATG5 with an affinity in the range of 10-20 nM (Figure 105 f-h). The controls showed no interaction between ATG5wt and GST (Figure 104 a) and no significant injection heat was detected, while injecting ATG16L1(11-43) peptide into pure buffer (Figure 104 b). Table 21 shows an overview of the resulting affinities.



	cell	syringe
a)	Atg5wt	GST
b)	buffer	Atg16L1(11-43)

Figure 104: ITC interaction controls. No binding was observed between ATG5wt and individual GST. Only small amounts of heat were produced by injecting ATG16L1(11-43) peptide into pure buffer (compare scales with Figure 105).



	cell Atg5	syringe Atg16L1	K_D [nM]
a)	wt	11-43	1050
b)	T75E	11-43	870
c)	T75A	11-43	1100
d)	wt	1-43 (GST)	250
e)	T75E	1-43 (GST)	320
f)	wt	1-30 (GST)	10
g)	T75E	1-30 (GST)	19
h)	T75A	1-30 (GST)	21

Figure 105: ITC interaction studies, using ATG5wt, T75E and T75A with ATG16L1(11-43), GST-ATG16L1(1-43) and GST-ATG16L1(1-30) as indicated in the table together with the resulting K_D values.

To confirm these results, the binding studies were repeated, using BLI as a complementary method. The His-tagged ATG5-constructs were coupled to anti-His tips. GST-ATG16L1(1-30) interacted specifically with all ATG5 variants with a high affinity of 42-50 nM (Figure 106, Table 21). Every experiment was run in triplets to obtain statistical significance. GST-ATG16L1(1-43) interacted unspecifically with the tips and could therefore not be used in BLI experiments. The synthetic peptide

ATG16L1(11-43) wasn't used neither, as BLI was sensitive for size differences and the MW of the peptide was only 4.3 kDa, which corresponded to only 12% of the MW of ATG5 (compared to 87-91% for the GST fusion proteins). In a control experiment individual GST didn't bind to the tips or to any of the ATG5 variants.

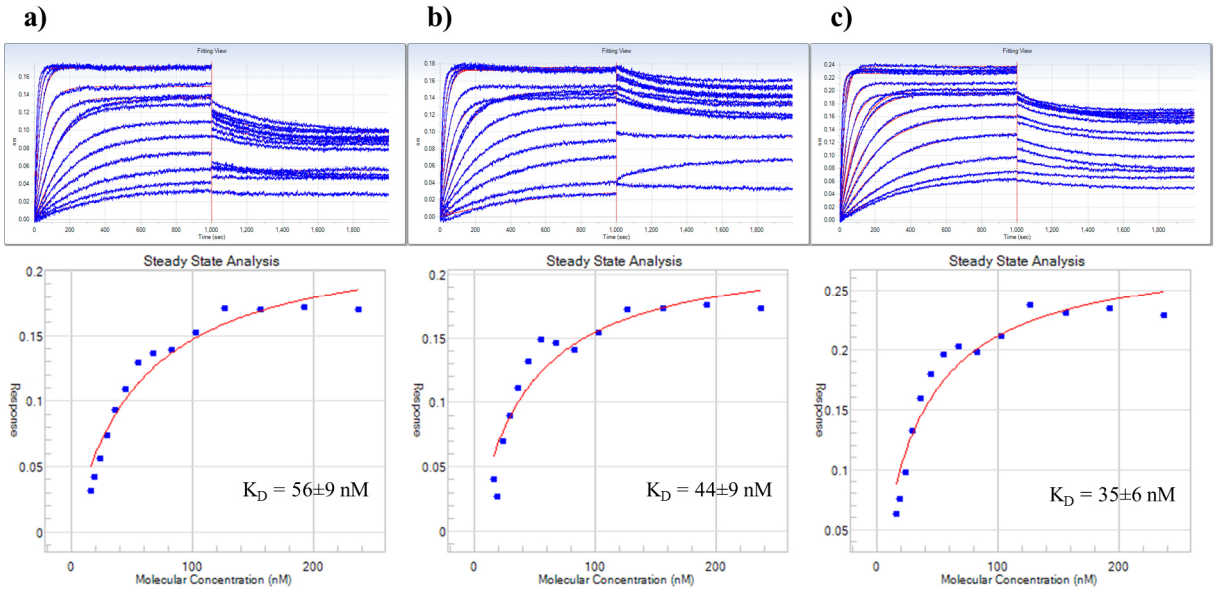


Figure 106: BLI interaction studies, using a) ATG5wt, b) T75E and c) T75A with GST-ATG16L1(1-30). Every experiment was run in triplets, but only the results of a single experiment are shown here. The K_D-values of the individual experiments are indicated in the graphs, the averaged K_D-values are indicated in Table 21.

5 Discussion

5.1 VirG

One objective of this work was the structural characterization of *Shigella* VirG, the outer membrane protein that is responsible for ABM. To set up crystallization experiments, milligram amounts of this protein were necessary. For this reason test-expressions of initially eight different constructs (Figure 28) were performed to identify optimal expression conditions. They were expressed as Strep- and as GST-tagged fusion proteins (Figure 29). Also coexpression of the binding partners ATG5 and IcsB was tried (Figure 30) as well as periplasmic expression (Figure 31), but in all cases only low protein yields were obtained. In large scale expressions the soluble protein proved to form aggregates, not suitable for crystallization experiments (Figure 32-Figure 35).

For this reason the ESPRIT system was used to screen for soluble constructs of VirG. 18 interesting constructs, covering regions that were not structurally characterized before, were identified, 13 containing the IcsB/ATG5 binding domain at least partially (Figure 36-Figure 39, Table 15). In Figure 107 the distribution of the N-termini of all VirG-constructs, resulting from ESPRIT screening, is shown. The biggest cluster of constructs contained N-termini between residues 590-640, corresponding to the autochaperone domain. Another cluster of constructs contained N-termini between residues 340-380, indicating that proteins with these N-termini were also stable. Finally a number of constructs didn't cluster, but were evenly distributed between residues 300 and 670. No constructs with N-termini before residue 340 were identified, except for the two constructs starting at residue 77 and 303, which carried a frame shift mutation in the beginning. Consequently constructs, extending further to the N-terminus than residue 340, were most probably unstable.

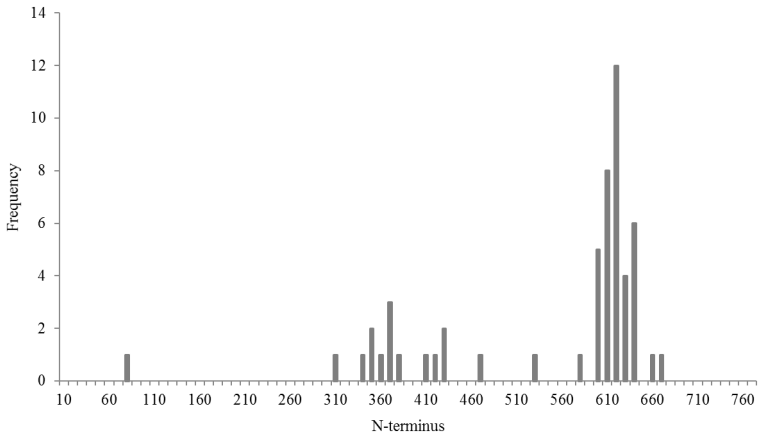


Figure 107: Histogram of the distribution of N-terminal residues in the VirG-constructs, obtained by ESPRIT. Many constructs began between residues 590-640, corresponding to the autochaperone domain. Another cluster of constructs began between residues 340-380.

Three of the ESPRIT-constructs, VirGE57, E94 and E76 were expressed and purified, which proved to be soluble and homogenous on SEC (Figure 40-Figure 43). VirGE94 crystallized (Figure 44) and the structure of this construct was solved (Figure 45).

5.1.1 VirGE94 structure

The structure of VirGE94 was composed of the C-terminal autochaperone domain, comprising residues 591-740 (blue in Figure 45), an intermediate part, comprising residues 484-590, that extended the β-helix by about four turns (green in Figure 45) and an N-terminal region, comprising residues 422-483, that formed the VID (orange in Figure 45).

The autochaperone domain was present in a conformation that was very similar to the published structure of the autochaperone domain (Figure 108 a). RMSD-Values were 0.45 Å for all atoms and 0.37 Å for the backbone residues. The major difference between the two structures was the conformation of an N-terminal loop, marked by an arrow in Figure 108 a. In the previously published structure of the isolated autochaperone domain the N-terminus was involved in a number of crystal contacts (Figure 108 b). The N-terminal loop itself was not involved, still these close by contacts probably influenced its conformation. In the VirGE94 crystal structure this loop did not contact symmetry related molecules, so the present conformation of the loop was likely to be the result of long range interactions and of the continuation of the β-helix.

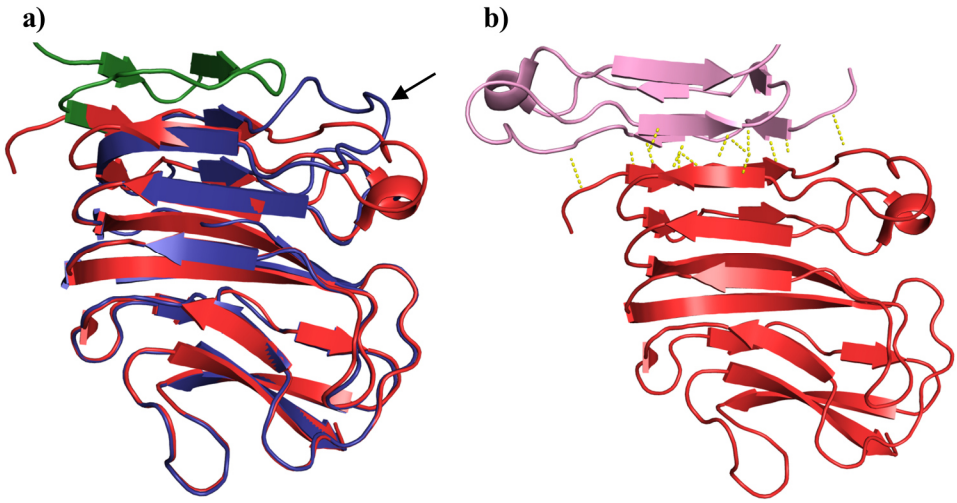


Figure 108: a) Comparison of the autochaperone domains of VirGE94 (blue) and the previously published structure from PDB-entry 3ML3 (red). The major difference was the conformation of an N-terminal loop (marked by a black arrow). The continuation of the β-helix in VirGE94 is shown in green. b) The N-terminus of the 3ML3 structure was involved in a number of crystal contacts (yellow dotted lines) with a symmetry related molecule (pink) that were likely to influence the conformation of this loop.

The most N-terminal part of VirGE94, comprising residues 422-483, formed a previously unknown domain, which we termed the VID and which consisted of two opposing β-sheets and a short α-helix in a connecting loop. The first β-sheet comprised two short parallel β-strands, while the second β-sheet comprised three antiparallel β-strands. A magnification of the structure of the VID and a schematic view on its fold is shown in Figure 109.

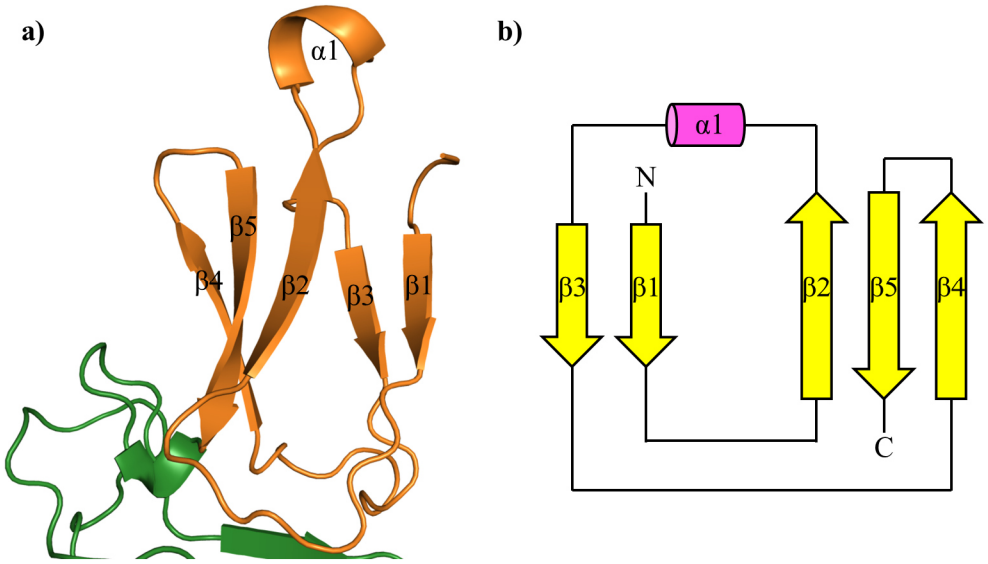


Figure 109: a) The structure of the VID, which consisted of two opposing β-sheets, the first comprising two short parallel β-strands, the second comprising three antiparallel β-strands, and a short α-helix in a connecting loop. b) A schematic view on the fold of the VID. The secondary structure elements are numbered in both panels accordingly.

The VID was located on top of the β-helix. The interface between the VID and the β-helix was mainly formed by hydrophobic residues, namely W432, L460, L461, I466, Y468, L483 and W485 on the side of the VID, as well as F496, L498, L504, V506 and L510 on the β-helical side (Figure 110). Consequently the interaction between the β-helix and the VID in the observed fashion, shielded the hydrophobic residues on both sides from the contact with the solvent.

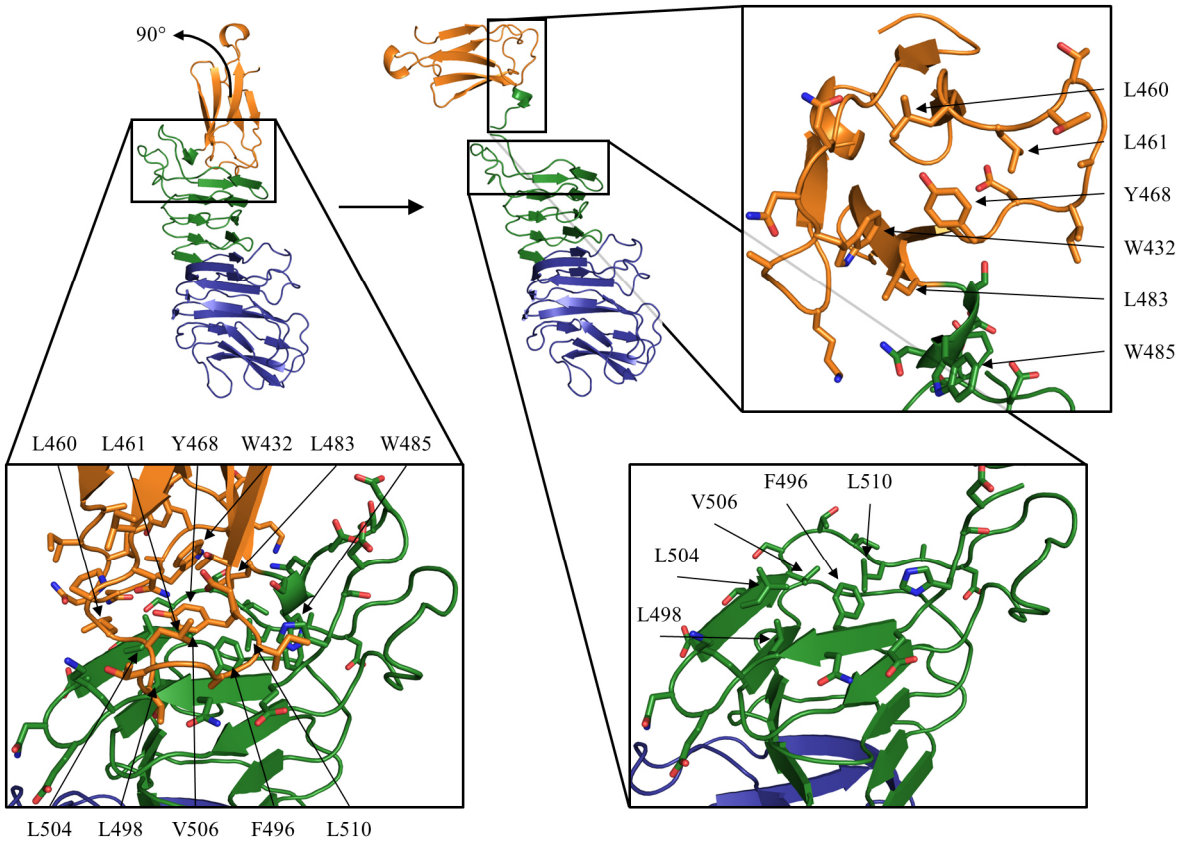


Figure 110: On the left the interface between the VID (orange) and the β-helix (green) is shown, which was mainly formed by hydrophobic residues. On the right the interface was opened by a 90° turn of the VID and both faces are shown for clarity.

A loop comprising residues 514-521 was flexible and its electron density was poorly defined, especially in chain B. In chain A the density was slightly more pronounced, probably due to the different surroundings, given the fact that in both cases the loop was not involved in crystal contacts. For either loop the backbone was traced but several side chains were completely invisible. Residue 758 was the C-terminus of VirGE94, but in the crystal structure only the amino acids up to residue 740 were visible. This indicated that the last 18 amino acids were flexible and might have formed a linker that connected the extracellular α-domain to the β-barrel membrane anchor.

A negatively charged cleft was formed by the four asparte residues D487, D489, D491 and D523, which were located in two extended loop regions at the N-terminus of the β-helix ((2) in Figure 111). This cleft might have had a function in the interaction with other compounds or proteins. It was clearly distinct from the polar localization region, which was located in the β-helical part of the structure below, where it comprised about two β-helical turns. In this region a negatively charged cleft was located that was formed between extended loops of the β-helix ((1) in Figure 111). The negative charge resulted from the three glutamate residues E499, E502 and E532. This cleft might be involved in the binding of an unknown factor that leads to the unipolar localization of VirG in the bacterial periplasm. The VID showed a mixed surface charge. A strongly positively charged patch ((3) in Figure 111) was formed by the three lysines K424, K438 and K478 and might be important for the function of the VID.

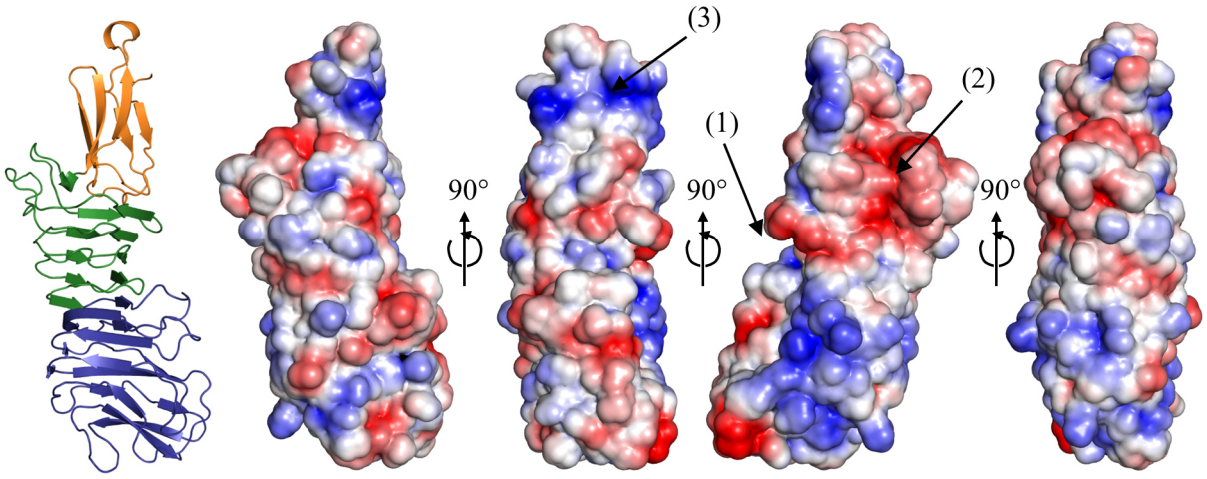


Figure 111: VirGE94 electrostatic surface potential, calculated, using APBS [199] [200]. Potentials between -3 kT/e (red) and +3 kT/e (blue) are color coded. Prominent features were a negatively charged cleft in the polar localization region (1), a second negatively charged cleft (2) and a positively charged patch in the VID (3). For orientation on the left a cartoon representation of VirGE94 according to the leftmost surface potential is shown.

The structures of a number of autotransporter proteins, different from VirG, had previously been published, including adhesins, serine proteases and a pore forming exo-toxin (Figure 112). They all had a central β-helix in common, which mainly acted as a rigid scaffold. Additional domains were attached to this scaffold in a modular fashion that exerted different protein specific functions. The *Bordetella pertussis* adhesin P.69 pertactin contained a long flexible loop that mediated the interaction with epithelial cells. In all five serine protease structures a large domain, containing the protease function, was present, which was attached next to the N-terminus of the β-helical stem, as well as a number of other domains of unknown function, designated domains 2 to 4. The *Helicobacter pylori* pore forming exo-toxin VacA contained an unusual domain, extending its C-terminus (Figure 112).

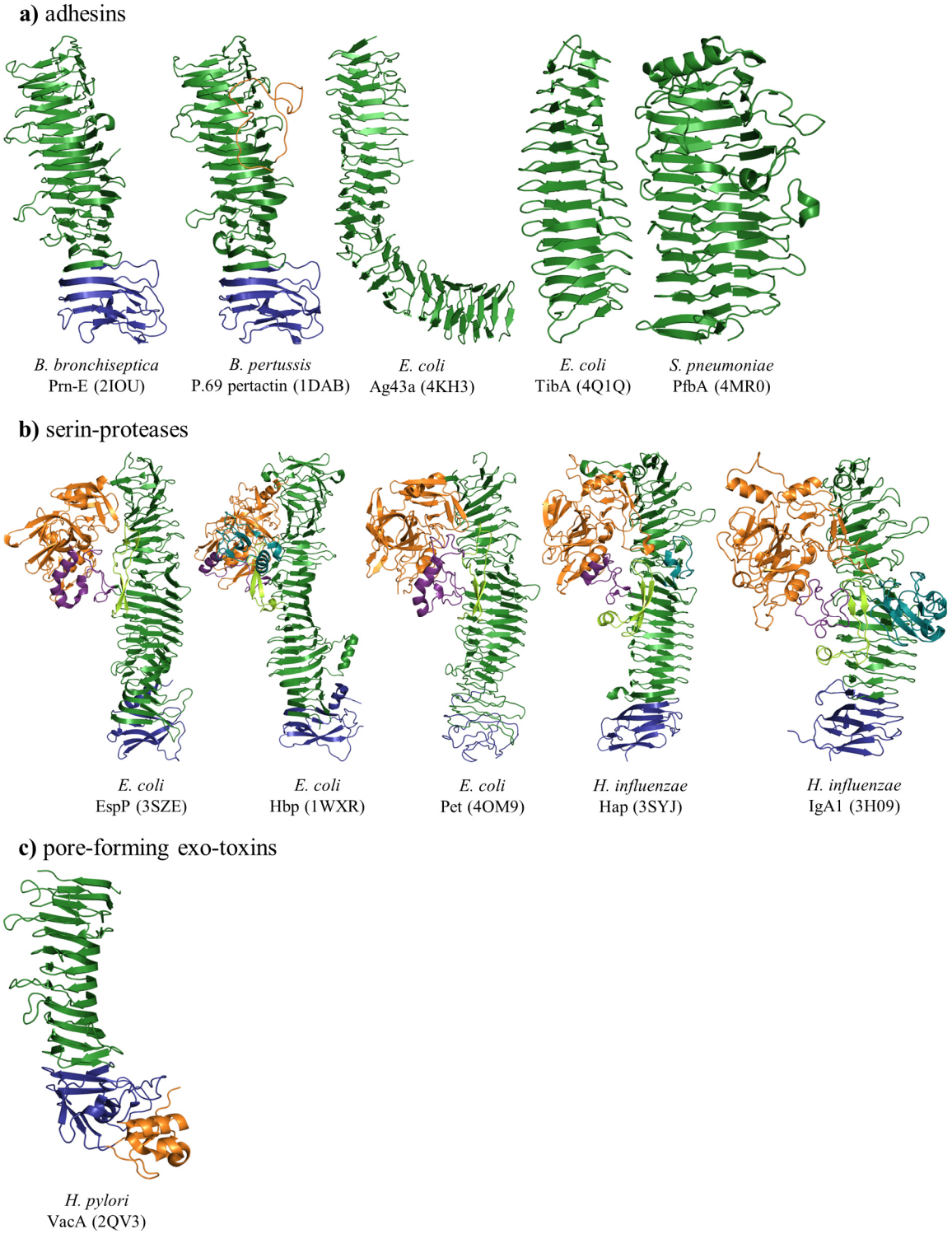


Figure 112: Overview of the known structures of different autotransporter proteins. For each structure the organism, the protein name and the PDB accession code are indicated. The β-helical stem (green), the autochaperone domain (blue) and different functional domains are highlighted in colors, where appropriate. In the structure of the adhesin P.69 pertactin, the orange loop contains the binding motif. In the serine-protease structures domain 1 (orange, the protease-domain), domain 2 (cyan), domain 3 (purple) and domain 4 (yellow) are highlighted. The structures are aligned, so that the N-termini are located at the top and the C-termini at the bottom. The membrane anchor domains, which insert into the outer bacterial membrane, follow at the C-terminus, hence they would be located at the bottom of these structures.

Compared to these known autotransporter structures, the autochaperone domain and the β-helical part of the VirGE94 structure were not unusual, as similar motifs were present in nearly all structures. However, in none of these structures a domain sitting on top of the β-helical backbone had been observed. Additional domains were rather attached to the side of the β-helical stem. As the VirGE94

construct did not cover the complete VirG α-domain, this raised the question, if the VID normally was located on top of the β-helix, as seen in the VirGE94 structure, or if it was attached to the side of an extended β-helical backbone, as seen in the structures of many other autotransporter proteins. This question is addressed in chapter 5.1.3.

5.1.2 VirGE94 oligomerization state

VirGE94 was dimeric in solution as shown by SEC (Figure 41). Two different crystal contacts were identified that might have been responsible for the dimerization. The first of these contacts was mediated by a heavy metal ion. As there were no heavy metals present in the buffers during purification or in the crystallization condition, the first possibility was that it was copurified from *E. coli* and the second possibility was that VirG had stripped off a Ni^{2+}-ion during His-affinity purification. When a Ni^{2+}-ion was placed into the density, it fitted perfectly, not leaving any unexplained electron density (Figure 113 a). It contacted residues H669 and D701 in both VirGE94 molecules. The aspartates contacted and neutralized the charge of the Ni^{2+}-ion, while the histidines complexed it. Consequently it was four times coordinated with distances to the aspartate δ-oxygens of 1.91/2.03 Å and to the histidine ε-nitrogens of 1.99/2.04 Å, which was quite close but had been observed before in other structures [201] (Figure 113 b). The B-factors of the Ni^{2+}-ion and the surrounding amino acids were of the same magnitude in a range between 20 and 28 $Å^2$, supporting the placement of the Ni^{2+}-ion. However, when the Ni^{2+}-ion was removed by dialysis with EDTA, VirGE94 still formed dimers in solution, as shown by SEC (Figure 46), indicating that this crystal contact was not responsible for the observed dimer formation.

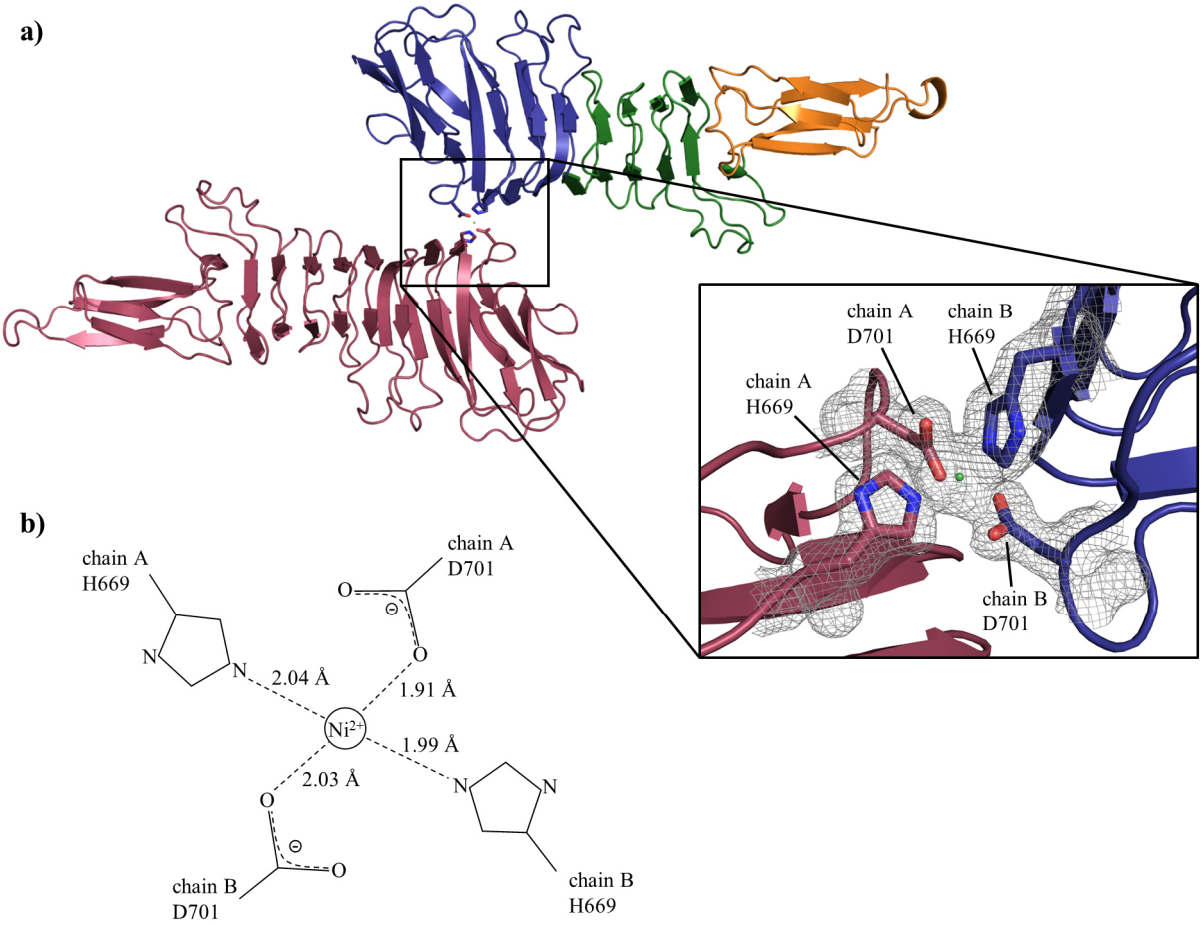


Figure 113: a) In the VirGE94 crystal structure a Ni^{2+}-ion (green sphere) was bound in between two symmetry mates of VirGE94 (one colored in blue, green and orange, the other in purple). It contacted residues H669 and D701 in both molecules. The inset shows a magnification of the contact area with a $2F_o$-F_c electron density map (grey) contoured at a level of 1.0 σ. b) A schematic view of the involved contacts, including the binding distances.

The second crystal contact that might have been responsible for VirGE94 dimerization, was an interaction of the VIDs from two symmetry related copies of VirGE94, which were rotated by 180° (Figure 114). The two VIDs formed an intermolecular β-sheet primarily by contacts between the two protein backbones. The only side chains that were involved in this interaction, were the side chains of S441 from both molecules.

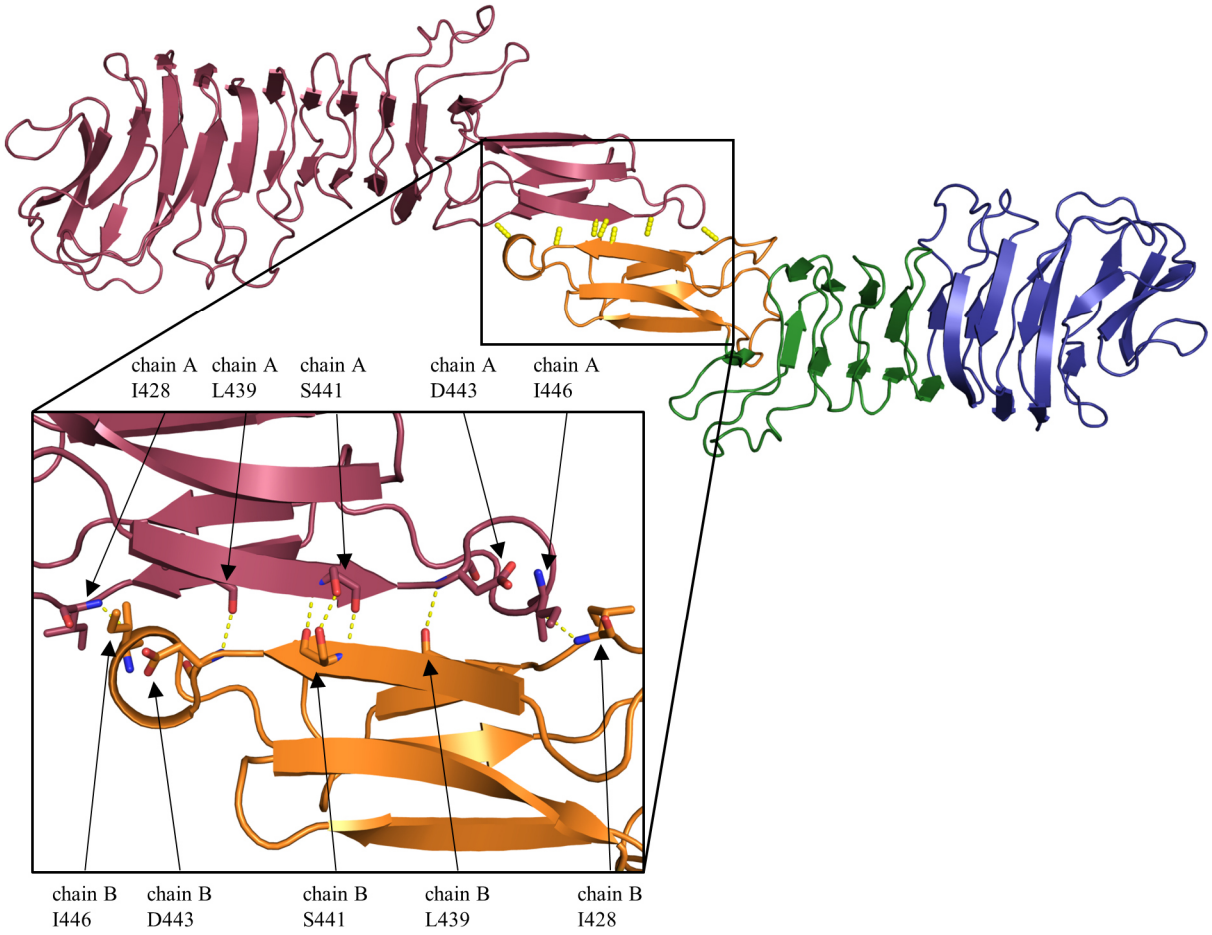


Figure 114: The VID (orange) contacted a symmetry related copy of itself, by the formation of an intermolecular β-sheet. The inset shows a detailed view on the participating residues.

To test whether this second crystal contact, mediated by the interaction of the VIDs, was responsible for the dimer formation of VirGE94 in solution, a SAXS experiment was performed (Figure 48). The MW of VirGE94 was calculated from the SAXS data and confirmed that VirGE94 was dimeric in solution (Table 17). The theoretical scattering curves of all possible crystallographic dimers were calculated and fitted to the experimental data. The dimer, formed by the interaction of the VIDs, resulted in the best fit (Figure 49) and also fitted into the envelope, calculated from the SAXS data, which had a similar elongated form (Figure 50 c). Also the results from the PISA server [202] suggested, that this dimer with an interface area of 679 Å^2 might have been stable in solution. Together these results confirmed that the VirGE94 dimer was formed by the interaction of the VIDs.

However, it was unlikely that the dimerization *via* the VIDs had an *in vivo* function. Normally the VirG proteins were located on the surface of the bacterial outer membrane, oriented in the same direction. Thus the formation of dimers by an interaction of the VIDs would have required a rotational freedom of about 90° to arrange the VIDs in a fashion, suitable for interaction (Figure 115). Additionally the longer VirgE76 construct was monomeric in solution, as shown by SEC (Figure 41), indicating that the dimerization interface might not have been accessible in longer VirG-constructs and in the full length

VirG α-domain. Together these results indicated that the dimerization *via* the VIDs was an artifact of the VirGE94 construct.

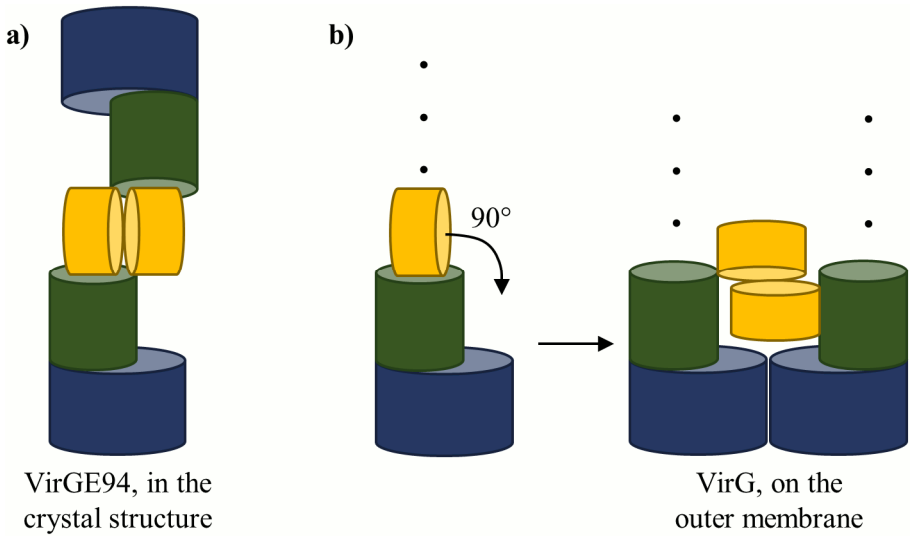


Figure 115: a) Scheme of the VirE94 dimer in solution. b) For the dimerization of VirG on the outer bacterial membrane, the VIDs would have required a rotational freedom of about 90°. The continuation of the VirG α-domain is indicated by black dots.

5.1.3 Localization of the VID

In contrast to many autotransporter proteins, where functional domains were attached to the side of an elongated β-helical stem (compare Figure 112), in the crystal structure of VirGE94, the VID was located on top of the β-helix. However, VirGE94 lacked the N-terminal half of the α-domain that might have formed a continuation of the β-helix. This raised the question of the position of the VID in the full length protein. Different scenarios were imaginable (Figure 116). The VID might have been attached to the side of a continued β-helical stem, like in other autotransporter proteins. In this case, in the truncated VirGE94 it might have moved to the unoccupied space on the top of the β-helix, to shield its hydrophobic interior from the solvent (Figure 116 a). Other possible locations of the VID in full length VirG were on top of the β-helix, like in the VirGE94 crystal structure (Figure 116 b), or it might have been aligned with the β-helix, to form a continued β-helical stem (Figure 116 c).

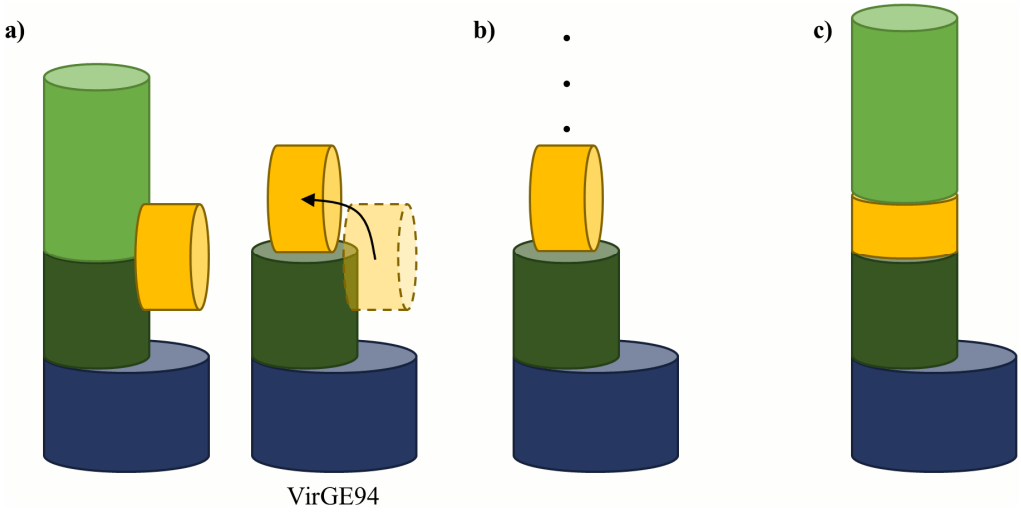


Figure 116: Models for the localization of the VID in the full length VirG α-domain. a) In full length VirG the VID (orange) might have been attached to the side of an extended β-helical stem (dark and light green). In VirGE94 the N-terminal half of the β-helix (light green) was truncated, hence the VID could have moved to the top of the C-terminal part of the β-helix (dark green) to shield its hydrophobic interior. At the bottom the autochaperone domain is shown (blue). b) The VID might have been located on top of the β-helix, like in the VirGE94 crystal structure, and the β-helix continues elsewhere (indicated by the black dots). c) The VID might have been aligned with the β-helix to form a continued β-helical stem.

A structure prediction of the VirG passenger domain by the Phyre2 server [197] showed a long β-helical stem and two domains attached to its side (orange and cyan in Figure 117). One of the domains in the Phyre2 prediction (orange in Figure 117) had residue 420 as N-terminus, which was nearly identical to the N-terminal boundary of the VID. In the prediction, this domain was composed of two separate chains (residues 420-456 and 497-534) that extruded from different loops in the β-helix, while in the crystal structure the VID was composed of a single chain. Also limited proteolysis experiments suggested that residues 420 and 485 might have been the boundaries of a domain in VirG (Figure 53). These indices confirmed the presence of the previously unknown domain in VirG. They also suggested that the VID corresponded to the orange domain in the Phyre2 prediction and might have been located at the side of an extended β-helix.

Figure 117: Phyre2 prediction of the structure of the passenger domain of VirG [197]. Models, used by Phyre2, were the structure of the VirG autochaperone domain (PDB-entry 3ML3) and the published structures of different autotransporter proteins (PDB-entries 1DAB, 3AK5, 3H09, 3SYJ, 4OM9). Blue: autochaperone domain, green: β-helical stem, yellow: N-terminal signal peptide for sec-pathway, orange and cyan: two domains attached to the side of the β-helical stem.

The question was addressed experimentally by a SAXS experiment, using VirGE57, E76, E94, K26 and K34. The expected solution states of the constructs are shown in Figure 47 b. To locate the VID in these constructs, their envelopes were calculated from the experimental scattering data (Figure 50). The envelope of VirGE76 was of special interest. It better resembled the full length VirG α-domain than VirGE94, as it contained the VID probably together with an extended β-helix.

VirGE94 was dimeric and a model of a VirGE94 dimer, formed by the interaction of the VIDs, fitted perfectly into the calculated envelope (Figure 50 c). The VID was located on top of the β-helix as in the crystal structure. For VirGE57 a truncated model of VirGE94 was fitted into the envelope (Figure 50 a). In this construct the VID was not complete, hence it was probably unfolded and its conformation and position were not significant. In VirGE34 the VID was deleted to test if further N-terminal regions of VirG would fold into a continued β-helix. A model of VirGE94, where the VID was deleted and which lacked the additional N-terminal amino acids, was fitted into the envelope (Figure 50 b). The unexplained volume at the N-terminus might have been attributed to the amino acids missing from the model, but it was not possible to determine their conformation. As the VID was deleted, no conclusions about its position could be drawn from the data of this construct. Unluckily the most important construct VirGE76 was aggregated and no envelope could be calculated. Also the other constructs showed signs

of aggregation first in their distance distribution functions (compare Figure 48) and second in the form of tails, extruding from the envelopes (compare Figure 50).

Consequently the question of the position of the VID in full length VirG could not finally be answered, as the SAXS data for the VirGE76 construct was missing. All scenarios were still possible.

5.1.4 Function of the VID

The structure of the VID was employed for the search of similar protein structures, using the DALI server [203] and PDBeFold [204]. A number of domains with similar folds were identified by the DALI server. Surprisingly the fold of the VID was similar to the fold of the autochaperone domains of different autotransporter proteins (Figure 119).

When the fold of the VID was compared to other folds associated with the β-helical backbones of known autotransporter proteins, it was clearly similar to the autochaperone domain fold, but distinct from all other folds (Figure 118).

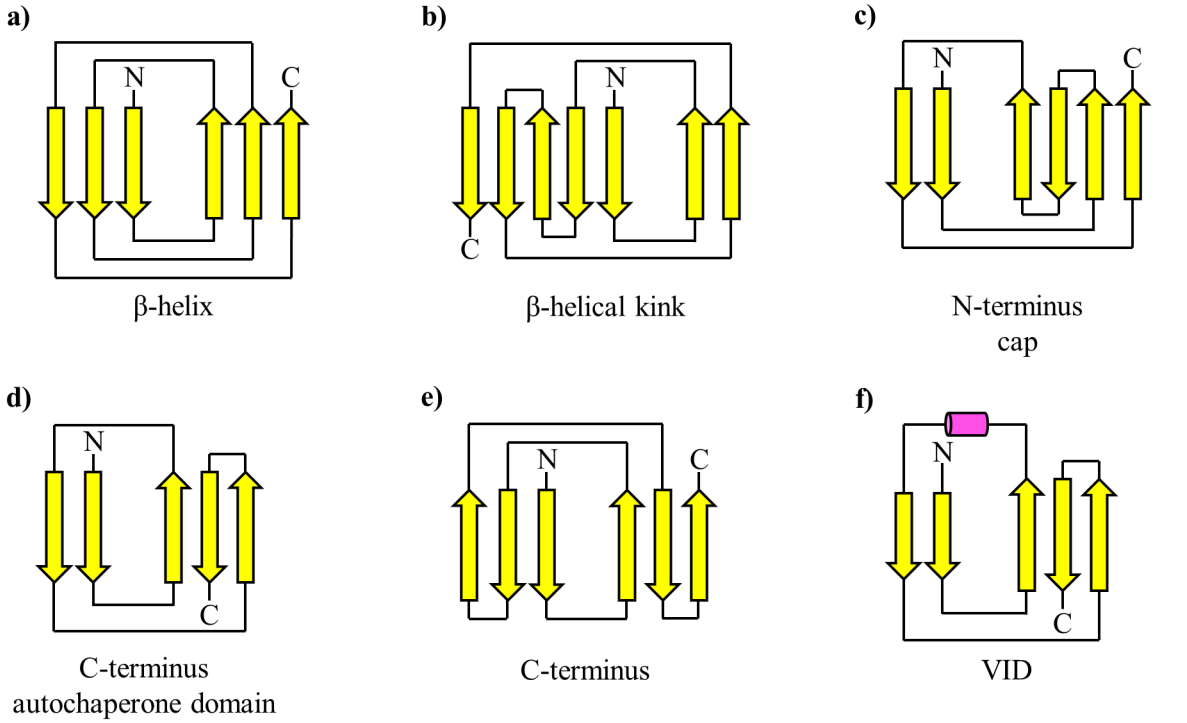


Figure 118: Different folds, associated with the β-helical structures of autotransporters. a) parallel β-helix, b) β-helical kink (PDB-entry 4KH3), c) N-terminal cap (PDB-entries 1WXR, 3H09, 3SYJ, 3SZE), d) C-terminal autochaperone domain (PDB-entries 1DAB, 1WXR, 2IOU, 2QV3, 3H09, 3SYJ, 3SZE, 4OM9), e) C-terminus (PDB-entry 4MR0), f) VID of VirGE94. The fold of the VID is clearly similar to the fold of the autochaperone domain, but distinct from all other folds.

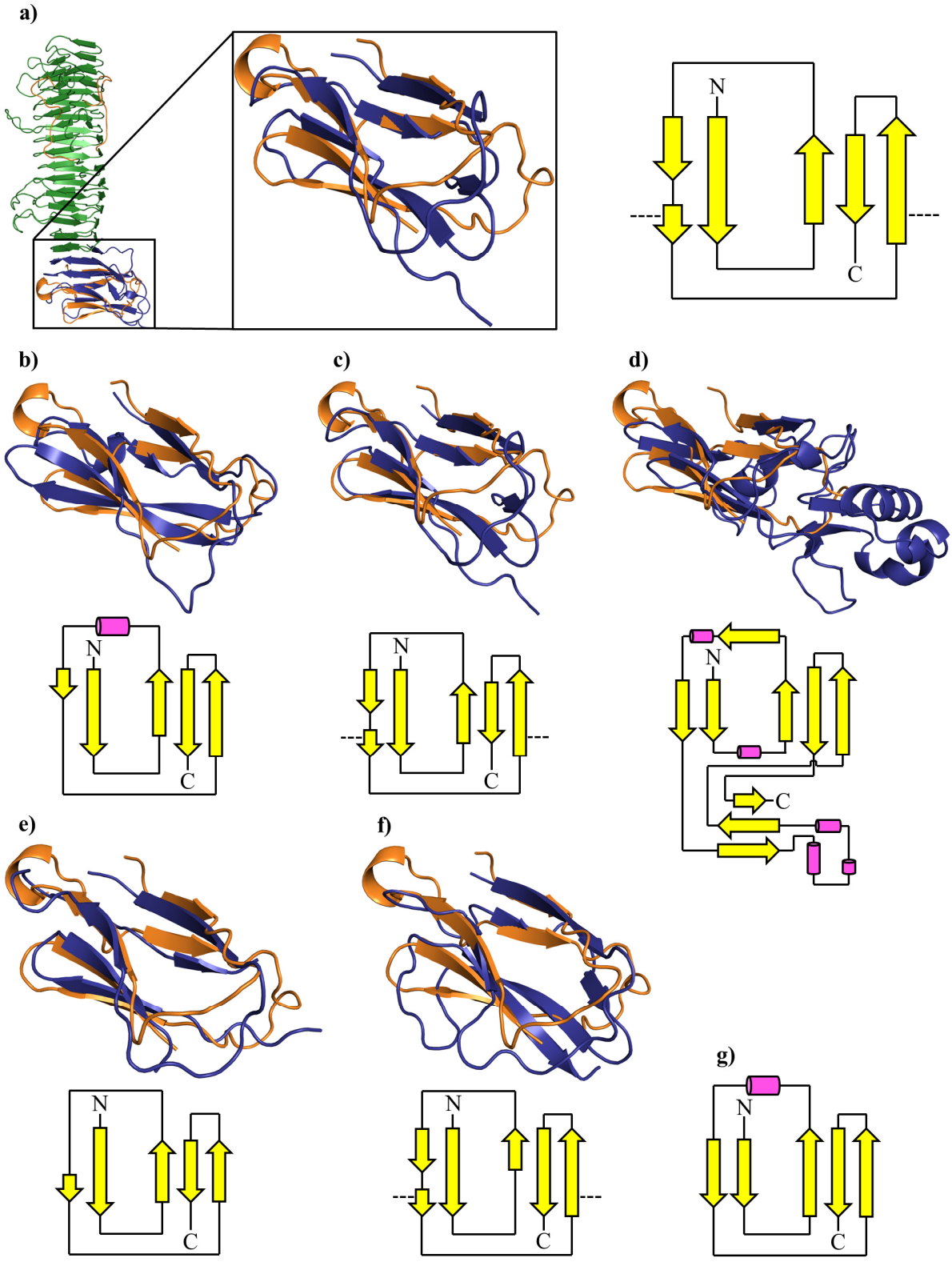


Figure 119: a) The best hit for a protein with similar fold, identified using the DALI server, was the autochaperone domain of *B. pertussis* P.69 pertactin (PDB-entry 1DAB) with an RMSD-value of 2.7. An overlay of the VID with the whole protein structure is shown. The inset shows a magnification of the VID (orange) and the P.69 autochaperone domain (blue). A scheme of the fold of the P.69 autochaperone domain is shown on the right, which is similar to the fold of the VID (compare with g)). A number of other autochaperone domains have also been identified by the DALI server, including the autochaperone domains of b) *E. coli* Hbp (PDB-entry 1WXR), c) *B. bronchiseptica* Prn-E (PDB-entry 2IOU), d) *H. pylori* VacA (PDB-entry 2QV3) and e) *H. influenza* IgA1 (PDB-entry 3H09). Overlays of the structures of the autochaperone domains (blue) with the VID (orange) are shown for all structures, above a scheme of the fold of the according autochaperone domain. All folds were similar to the fold of the VID (compare with g)), even if in the case of VacA an additional domain and different secondary structure elements were inserted. f) The VirG autochaperone domain had not been identified by the DALI server, still a structural alignment and the scheme of the fold showed the similarity to the VID. g) Scheme of the fold of the VID the VirGE94 structure.

Conclusively the VID might have acted as a second autochaperone domain, which nucleated the fold of the N-terminal continuation of the β-helix. Consequently the interface, which formed the dimer in VirGE94, would not have been accessible anymore in N-terminally elongated VirG-constructs and in the full length VirG α-domain. The fact that only VirGE94 dimerized, but the shorter VirGE57 and also the longer VirGE76 were monomeric, supported this hypothesis (Figure 41). An autochaperone domain function of the VID created new possibilities for the conformation of the VirG α-domain (Figure 120). One scenario was that the VID might have been located as in the VirGE94 crystal structure, and introduced a 90° kink into the protein backbone (Figure 120 a). It might also have been located to the side of a continued β-helical backbone and nucleate the fold of a separate β-helical domain there (Figure 120 b). A last possibility was the formation of a continued β-helix, similar to other autotransporter proteins (Figure 120 c). A kink in the β-helical backbone would not have been too extraordinary, as a nearly 90° kink, which was formed by other means, had already been observed in the Ag43a crystal structure (Figure 112).

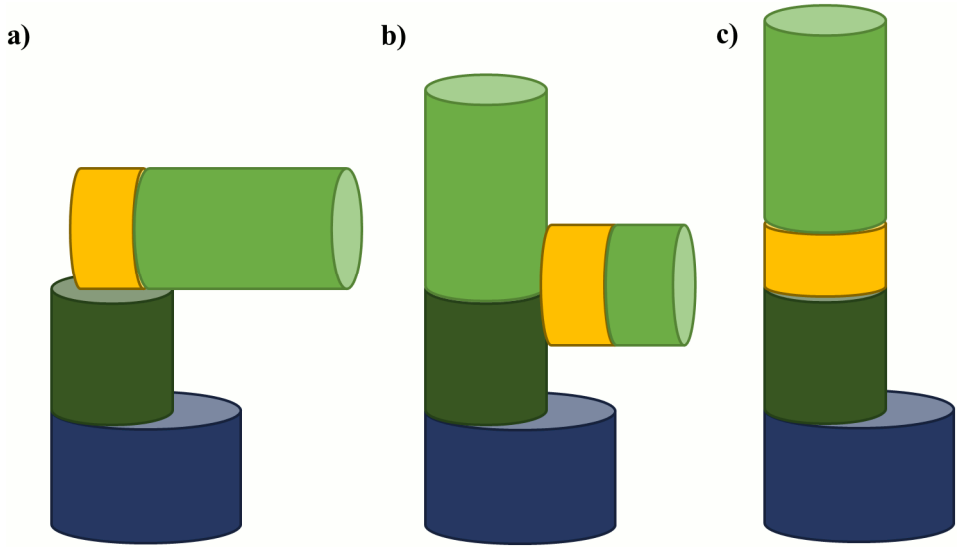


Figure 120: Schemes of possible conformations of the VirG α-domain. a) If the VID was located, like in the VirGE94 crystal structure, it might have introduced a 90° kink in the β-helix. b) If the VID was located at the side of a continued β-helical stem, it might have nucleated the formation of a separate β-helical domain there. c) The VID might have formed a continuation of the β-helix.

5.1.5 Interaction studies

The previously published interactions of VirG with IcsB and ATG5, which were the basis of the *Shigella* camouflage mechanism, were evaluated here, using various protein constructs and methods, for a more detailed understanding. The interactions of the initial VirG-constructs were assessed in pulldown experiments (Figure 88), by blue native PAGE (Figure 89) and by SEC (Figure 90), but binding of IcsB or ATG5 was not observed. The major problems were the low available amounts of VirG protein and the fact that it was aggregated, as shown by SEC.

For this reason the interaction of VirG with IcsB and ATG5 was reassessed, using the soluble and homogenous VirGE94 and E76 constructs, which resulted from the ESPRIT screening. Still in pulldown experiments no interactions with IcsB or ATG5 were observed (Figure 91), but using MST as a complementary method an interaction of VirGE76 and IcsBwt (in complex with IpgA) with a K_D of 600±205 nM, but not of VirGE94 and IcsBwt, was observed (Figure 92). This was not surprising as VirGE94 contained only the last 15 amino acids of the IcsB/ATG5 binding domain, while VirGE76 contained nearly the full domain, except for the first 20 amino acids. However, it was strange that an interaction with a K_D of 600 nM was not observable in the pulldown interaction studies. The reason might have been that VirG was untagged and pulldowns could only be performed, using the His-tag of

IcsB. By binding IcsB to the affinity beads, the VirG binding site might have been blocked. Still the interaction between VirGE76 and IcsB should be considered as uncertain, until being validated by an independent method. In the case of ATG5 no clear interaction to VirGE76 or VirGE94 was detected, indicating that either ATG5 did not directly bind to VirG or that the ATG5 binding site included residues 320-339 of the previously reported interaction motif, which were not included in the soluble constructs, identified by ESPRIT and used in these experiments.

VirG has been shown to be an activator of N-WASP in a fashion similar to Cdc42. Still it is quite unclear which domains of N-WASP and VirG are involved in their interaction. Binding of N-WASP-PH/IQ [194], -GBD [85] and -V [145] to VirG has been shown. At the same time three different interaction regions in VirG have been shown to bind to N-WASP and N-WASP interacts with different VirG molecules at the same time [135]. To further investigate the interaction between N-WASP and VirG, pulldowns (Figure 93) and MST experiments (Figure 94) were performed, but no interactions were observed. This did not mean that there was no interaction between VirG and N-WASP. Only the individual N-WASP domain has been used in these experiments and VirGE76 only covered N-WASP interaction regions III and II (except for the first ten amino acids), but not N-WASP interaction region I, which might have been important for this interaction. A VirG-construct covering N-WASP interaction regions I and II was the first one, for which *in vitro* interaction with N-WASP had been shown [146]. Hence the constructs, used in these experiments, might have been too limited to allow for the interaction of VirG and N-WASP.

Finally a pulldown experiment was performed to test the interaction of VirG to other proteins involved in or associated with the actin polymerization system. The tested interaction partners were domains of N-WASP, Toca-1, Cip4, Cdc42 and a number of diaphanous formins (Figure 95). No interactions between VirGE76 and these constructs were observed. However, as VirGE76 covered only a part of the VirG α-domain and only individual domains of the potential binding partners were used, an interaction between full length VirG and the respective proteins cannot be excluded here.

5.2 IcsB

IcsB was the second target that should be structurally characterized in this work. Initially twelve different constructs were used, which were expressed and purified in high quantities (Figure 59-Figure 62). The proteins were soluble and homogenous, but no initial crystallization conditions were identified. Different rescue strategies like SER (Figure 63-Figure 65), methylation of surface lysines and *in situ* proteolysis were applied, but didn't lead to crystal formation either. A limited proteolysis experiment was used to identify suitable boundaries for new constructs (Figure 66). Using these boundaries, 19 new constructs were designed (Figure 67). Proteins were expressed and purified, but again no initial crystallization conditions were identified (Figure 68, Figure 69). At this point attempts to crystallize IcsB were stopped and the purified proteins were used to characterize IcsB.

5.2.1 Biophysical characterization

The purified IcsB-IpgA complexes were homogenous and soluble, as shown by SEC. Using SourceQ IEC, IpgA was partially separated from IcsB. The resulting individual IcsB proteins, which contained the CBD, were unstable, as they aggregated, as shown by SEC. In contrast, individual IpgA remained soluble and homogenous. The reason for the lability of individual IcsB was the presence of the CBD. IcsB-constructs, lacking the CBD, were soluble and stable without IpgA. On the other hand, expression of the CBD (11 kDa) fused to GST as a solubility tag (28 kDa) without coexpressed IpgA, resulted in the complete aggregation of the fusion construct. The absence of expression for IcsB protein constructs, including the CBD, without coexpressed IpgA, was in compliance with these results. Taken together, these results indicated that the presence of the CBD destabilized the IcsB protein, while it was stabilized by the presence of IpgA.

Upon secretion by the TTSS, IpgA is stripped off, resulting in the export of IcsB, including the CBD, to the cytoplasm of human cells. Hence the CBD might have an *in vivo* function that needs further investigation. Possibly it is a protein interaction domain that is stabilized, upon binding of an interaction partner derived from the host.

Thermofluor-assays showed that the IcsB-IpgA complex had an intermediate melting temperature of 50°C, compared to the melting temperatures of individual IcsB of 37°C and the melting temperature of individual IpgA of 72°C (Figure 70, Figure 71). This result supported the stabilizing function of IpgA. A melting temperature, close to the human body temperature, might be important for the function of IcsB, by providing a high degree of flexibility to the protein, while it is stabilized by IpgA as long as it resides inside the bacterial cell, to prevent degradation.

CD-spectroscopy was used to determine the secondary structure composition of IcsB (Figure 72). Interestingly the fraction of random coil rose from 25% to 37% upon subtraction of the spectrum of IpgA from the spectrum of the IcsB-IpgA complex. This major fraction of random coil in IcsB, might be the reason for the bad crystallization behavior of all IcsB-constructs. These results were in agreement with the findings from [205], where IcsB was shown to contain natively unfolded regions by NMR and that it was no suitable target for crystallography.

5.2.2 Mapping of the IpgA binding site

Previously, IcsB residues 171-247 had been shown to contain the minimal CBD [131]. However, in this work N-terminal IcsB-constructs, comprising the first 125 residues, repeatedly copurified IpgA (Figure 62). This led to the hypothesis that the CBD was actually located in the N-terminal part of IcsB. Using the IcsB-constructs, generated in this work, the CBD was mapped in pulldown experiments (Figure 73). The minimal fragment of IcsB that copurified IpgA was IcsBK3, spanning residues 26-125.

The CBDs of several TTSS effector proteins were located in the N-terminal region of their respective target proteins [206]–[208]. Using SIB-MyHits (Pfam) [209], [210], a stretch in IcsB, comprising residues 33-115, was predicted to contain a SicP binding site [153]. SicP was the chaperone of SptP, a TTSS effector protein in *Salmonella typhimurium* [206]. IcsB 33-115 shared 27% sequence identity (51% similarity) with SptP 44-121, the SicP binding site in SptP. The chaperones IpgA and SicP shared 20% sequence identity (34% similarity). These significant sequence conservations supported the hypothesis.

A limited proteolysis experiment has been performed to identify suitable boundaries for new IcsB-constructs. The identified N- and C-terminal boundaries of the fragments, were marked in a PSIPRED secondary structure prediction of IcsB [211] (Figure 121). Most cleavage sites were close to the N- and C-terminus and in a region around a presumably unfolded stretch (predicted by DISOPRED2 [212]). Some single cleavage sites were located in the C-terminal part of IcsB. Interestingly, no cleavage site was identified in an N-terminal region spanning residues 10-175, indicating that this region was protected from cleavage, complying with the previous results.

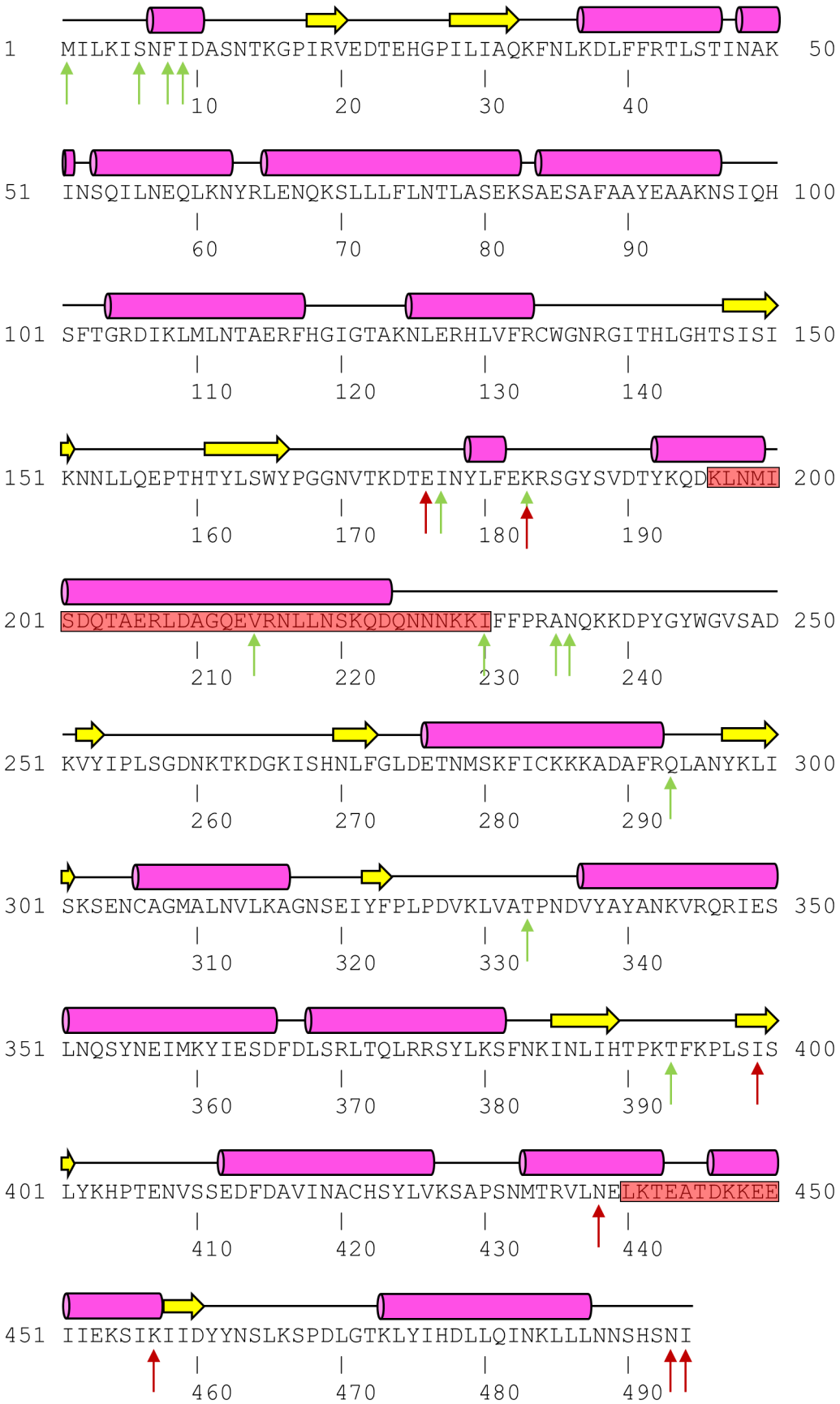


Figure 121: Mapping of the N- and C-terminal cleavage sites, determined in a limited proteolysis experiment, on a PSIPRED secondary structure prediction of IcsB. Green arrows indicate N-terminal fragment boundaries, red arrows indicate C-terminal boundaries. Red boxes indicate regions that were predicted as unfolded by DISOPRED2.

Finally a SAXS experiment was performed, using an IcsBK3-IpgA complex as well as individual IpgA (Figure 74). The SAXS data were used to determine the MW of the proteins (Table 18). The MW of IpgA corresponded to a dimer, while the MW of the IcsBK3-IpgA complex corresponded to one molecule of IcsBK3, binding a dimer of IpgA.

The structures of a number of TTSS effector proteins in complex with their respective chaperones were deposited in the PDB. An overview of these structures is given in Figure 122. Two models were created from each structure, one containing a chaperone dimer, the other containing a chaperone dimer bound to the CBD of their respective effector protein. These models were fitted to the experimental scattering curves. The χ^2-values of the resulting fits are shown in Figure 123. Fitting of the chaperone dimer models to the SAXS data of IpgA, consistently resulted in good fits with χ^2-values smaller than 4. If the same models were fitted to the SAXS data of the IcsBK3-IpgA complex, the χ^2-values in most cases were considerably higher, reaching values of 6-14. Fitting of the chaperone-effector complexes to the SAXS data of the IcsBK3-IpgA complex, resulted in much lower χ^2-values again, similar to those, obtained for the fit of the chaperone dimers to the SAXS data of IpgA. These results confirmed that IpgA formed a typical chaperone dimer, while IcsBK3-IpgA formed a chaperone-effector complex, where one molecule of IcsBK3 bound a dimer of IpgA.

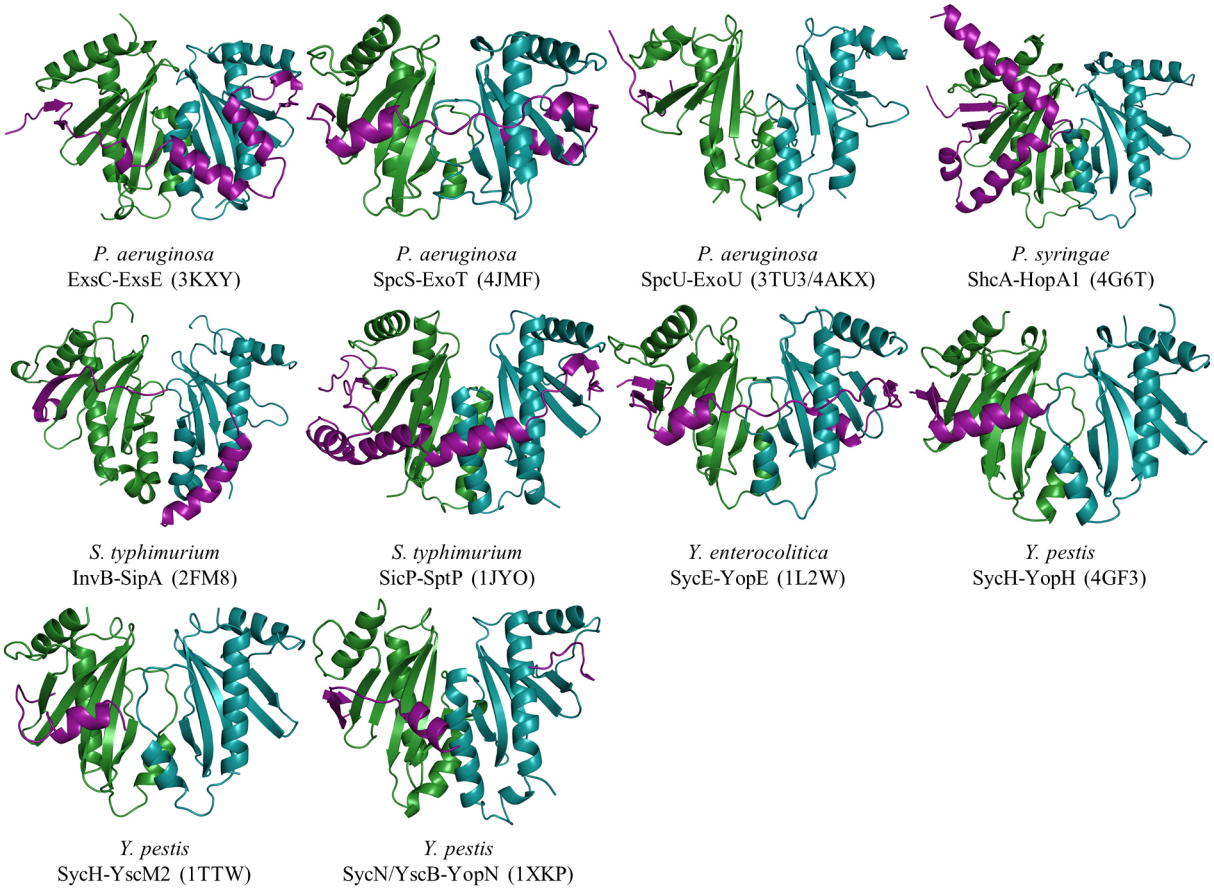


Figure 122: Overview of the known structures of different class I TTSS chaperone-effector complexes. For each structure the organism, the chaperone and effector names as well as the PDB accession code are indicated. The chaperone dimers are shown in green and blue, the CBD of the effector is shown in purple.

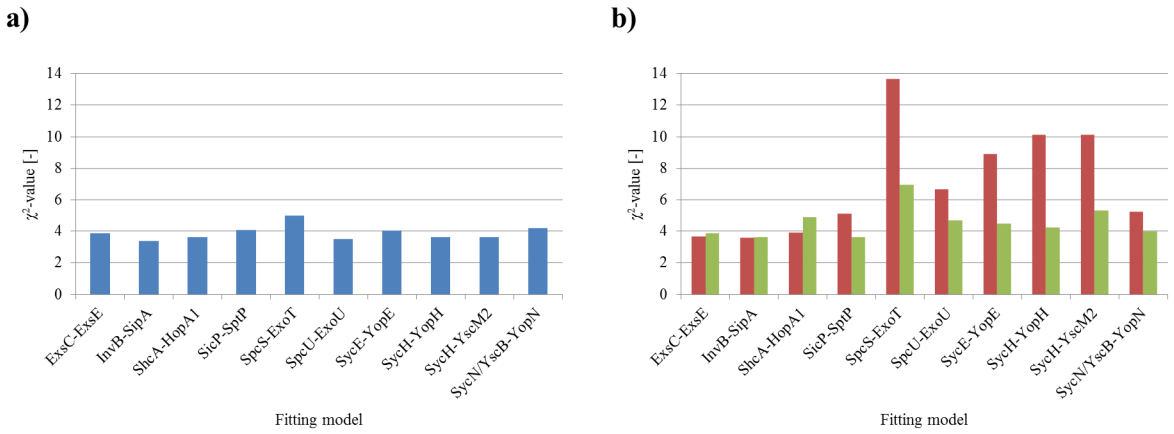


Figure 123: Fitting of different known TTSS chaperone-effector complexes to the experimental SAXS data of IpgA and of the IcsBK3-IpgA complex. a) blue bars: fitting of chaperone dimers to IpgA SAXS data, b) red bars: fitting of chaperone dimers to IcsBK3-IpgA SAXS data, green bars: fitting of chaperone effector complexes to IcsBK3-IpgA SAXS data. These results confirmed that IpgA formed a chaperone dimer in solution, while IcsBK3-IpgA formed a chaperone-effector complex.

Conclusively, the results presented here show that the IcsB-CBD is located in the N-terminal region, between residues 26-125, as shown by the pulldown experiments. SAXS experiments confirmed that an IpgA dimer was binding in this region, which is typical also in other TTSS chaperone-effector complexes.

5.2.3 Interaction studies

Interaction studies with published binding partners of IcsB were performed for the further characterization of IcsB. The reported interaction of IcsB with Toca-1 was tested in pulldown experiments, but could not be confirmed (Figure 96, Figure 97, Figure 103). In the original publication the interaction of IcsB and Toca-1 had been shown by immunoprecipitation from human cell lysates and Western blotting, and was confirmed by the coprecipitation of MBP-IcsB (no IpgA) and His-Toca-1(105-477) from *E. coli* lysates and Western blotting, which mostly excluded the possibility that a third interaction partner was involved. This leaves several possibilities, why the interaction could not be shown here. One reason might have been the presence of IpgA, which complexed IcsB. If the N-terminal part of IcsB was involved in the observed interaction, the presence of IpgA might have blocked it. Also in this work Toca-1(1-293) and (245-477) (the core IcsB binding domain), but not the longer construct (105-477) have been used, leaving the possibility that Toca-1 residues 105-244 were involved in the interaction. In the publication the interaction was shown by Western blotting, while here pulldown experiments and Coomassie stained gels were used. If the interaction was weak, it might have been disrupted by the washing steps, applied in the pulldowns, and might not have been detectable on Coomassie stained SDS-gels anymore. Also a major problem in the pulldown experiments was the strong and unspecific binding of GST-tagged Toca-1-constructs to His-affinity beads, which might have obscured an interaction.

Another published interaction partner of IcsB was cholesterol. This interaction was tested, using DHE, a fluorescent cholesterol analogon, in SEC. However, the interaction could not be confirmed. One problem might have been that DHE, like cholesterol, was virtually insoluble in water. For this reason methyl-β-cyclodextrin, a cyclic compound composed of seven glucose moieties, was used to solubilize DHE. If the solubilization didn't work, DHE might not have been available for interaction with IcsB. Another possibility was that the differences in the structure between cholesterol and DHE prevented the interaction.

A third potential interaction partner has been identified in this work. Phyre2 [197] was used to predict the tertiary structure of IcsB. The quality of the resulting model was poor, as no homologues structures

were known. Still, a stretch in IcsB, comprising residues 99-130, was shown to be homologues to a stretch in *E. coli* $EspF_U$, comprising residues 73-102. These two regions shared 30% identity (46% similarity). A more truncated core region, comprising residues 106-130 in IcsB and 75-99 in $EspF_U$, even shared 41% identity (63% similarity) (Figure 124).

```
EspF_U       75 DVAQRLMQHLAEHGIQPARNMAEHI 99
                *:   *      ***  *:*: :*:
IcsB        106 DIKLMLNTAERFHGIGTAKNLERHL 130
```

Figure 124: Alignment of IcsB(106-130) and $EspF_U$(75-99). These two regions shared 41% identity (63% similarity).

$EspF_U$ is an activator of N-WASP and a TTSS effector protein from entero hemorrhagic *E. coli* (EHEC). In its inactive state N-WASP is autoinhibited by an intramolecular interaction of the GBD with the C-region of the VCA-domain. $EspF_U$ binds competitively to the GBD, releasing the VCA-domain, which in turn activates the Arp2/3-complex. The direct interaction of $EspF_U$ with the WASP-GBD has been proven, by solving the solution structure of the complex by NMR (PDB-entry 2K42) [198]. $EspF_U$ is composed of an N-terminal signal peptide, followed by 2-7 nearly identical 47 amino acid repeats. Every repeat consists of a 27 aa N-terminal hydrophobic stretch and a 20 aa C-terminal proline rich region. The first 33 aa, which comprise the whole hydrophobic region and the first part of the proline rich region, are sufficient to effectively activate N-WASP [198]. The IcsB core region, identified above, corresponded to a stretch in $EspF_U$, covering all the hydrophobic part of one repeat, except for the first two amino acids. For this reason it was hypothized that IcsB might bind and activate N-WASP.

For this reason the two IcsB-constructs K30(104-130) and K31(99-136) were designed and tested together with IcsBwt (1-494, in complex with IpgA) and K9 (169-494, no IpgA) in pulldown experiments for interaction with the N-WASP-GBD (Figure 98, Figure 99, Figure 103). However, no interactions were observed. The presence of IpgA, complexed to IcsBwt, might have been problematic. The predicted N-WASP binding site (104-130) partially overlapped with the CBD (33-115). One solution might have been the coexpression of IcsBwt and the N-WASP-GBD, which has not been tested here. The constructs IcsBK30 and K31 also might have been too limited to support interaction with the N-WASP-GBD. In an independent approach N-WASP-GBD – IcsBK30/K31 fusion proteins were tested for interaction, using a thermofluor-assay (Figure 100, Figure 101). Here, for IcsBK31 and the N-WASP-GBD, indications for an interaction have been observed. Interestingly, in contrast to IcsBK30, IcsBK31 included IcsB residue W135, which was in a similar position as $EspF_U$ residue W105 that had been shown to contribute to a higher affinity N-WASP interaction [213]. Unfortunately, in this experiment the positive control did not work, hence the interaction between IcsB and the N-WASP-GBD should be considered as uncertain, until being validated by an independent method.

Like N-WASP, diaphanous formins contain a GBD and are autoinhibited by an intramolecular interaction. Diaph1 has been shown to be involved in the formation of *Shigella* protrusions [90]. As an interaction of IcsB and the N-WASP-GBD was hypothized, also the interactions of IcsB and the GBDs of different diaphanous formins were tested in pulldown experiments (Figure 102, Figure 103). Also here, no interactions were observed, possibly due to the weak soluble expression of most of the GBD-constructs.

Finally IcsB contained a predicted RID [153]. For this reason the interaction of IcsB and Cdc42 was tested in a pulldown experiment (Figure 103), but no interaction was observed.

In summary, a potential interaction of IcsB and the N-WASP-GBD has been shown, apart from the potential interaction of IcsB and VirG that was described in chapter 5.1.5. The presence of complexed IpgA in IcsBwt might have been problematic, as interactions with the N-terminal part of IcsB would

likely have been missed. Also in most cases, interactions were tested, using individual domains of the binding partners, which might not have been sufficient for interaction with IcsB.

5.3 ATG5

5.3.1 ATG5 structure

Initially only one ATG5-construct, ATG5wt, was expressed and purified (Figure 75, Figure 76). Subsequently ATG5wt was crystallized in November 2012 (Figure 78) and its structure was solved in April 2013 (Figure 79). At the beginning of this work, only two structures from December 2006 of *S. cerevisiae* Atg5 in complex with N-terminal fragments of Atg16 were known (PDB-entries 2DYM, 2DYO). Hence, the structural characterization of human ATG5 in this work was a logical objective. In August 2012 the structure of *Kluyveromyces marxianus* Atg5 was published (PDB-entry 3VQI), followed by the first human ATG5 structure in complex with ATG12 and an N-terminal fragment of ATG16L1 (PDB-entries 4GDK, 4GDL) and the homologues *S. cerevisiae* complex (PDB-entry 3W1S) in December 2012. At this point ATG5 has already been crystallized. The structure of individual ATG5 was interesting to evaluate possible conformational changes of the ATG5 mutants T75E and T75A, as well as in a potential VirG-ATG5 complex. In the published complex structure, conformational changes might have been imposed on ATG5. The structures of individual ATG5wt and the mutant forms were solved between April and December 2013. In November 2013 the structure of the human ATG5-12/16 complex, binding a fragment of ATG3 was reported (PDB-entry 4NAW) and only recently in March 2015 the structures of human ATG5 in complex an N-terminal fragment of ATG16L1 (PDB-entry 4TQ0) and the AFIM of TECPR1 were published (PDB-entry 4TQ1).

The structure of individual ATG5, solved in this work, is shown in Figure 79. Similar to *S. cerevisiae* Atg5 (Figure 15), human ATG5 consisted of two UFDs and an α-helical bundle region. The first and the last three amino acids were not modelled, due to missing electron density. Additionally, an internal loop comprising residues 228-234 was not modelled for the same reason. Most probably these residues were flexible and therefore not visible in the electron density map.

The peptide bond between residues L96 and P97 was in the cis-form, which is not unusual for a peptide bond, involving a proline residue. The resolution of 3.1 Å was not sufficient to determine the correctness of the cis- or trans-configuration, but the cis-peptide was already present in the molecular replacement search model, which was determined at a resolution of 2.6 Å. A human ATG5 structure published in 2013 (PDB-entry 4NAW) with a resolution of 2.2 Å also showed the cis-peptide, which was stabilized by an octahedrally coordinated ion (Figure 125). This ion contacted the backbone oxygens of A95 and P97, as well as the side chain of N99, thereby stabilizing the cis-configuration. The other three coordination sites were occupied by water molecules. The nature of this ion was not further specified in the PDB-file or the according publication [214].

The structure of individual ATG5 was very similar to the structure of ATG5 in the published human ATG5-12/16 complex (PDB-entry 4GDK) [215] with RMSD-values of 0.41 Å for the backbone residues and 0.47 Å for all atoms. Hence no conformational changes of ATG5, upon complex formation were observed. Human ATG5 shared 24% identity (42% similarity) with *S. cerevisiae* Atg5 and 24% identity (40% similarity) with *K. marxianus* Atg5. The RMSD-values were 1.41 Å for the backbone residues and 1.59 Å for all atoms for *S. cerevisiae* (PDB-entry 2DYM) and 1.74 Å for the backbone residues and 1.78 Å for all atoms for *K. marxianus* (PDB-entry 3VQI). The structures of the different ATG5 homologues were closely related.

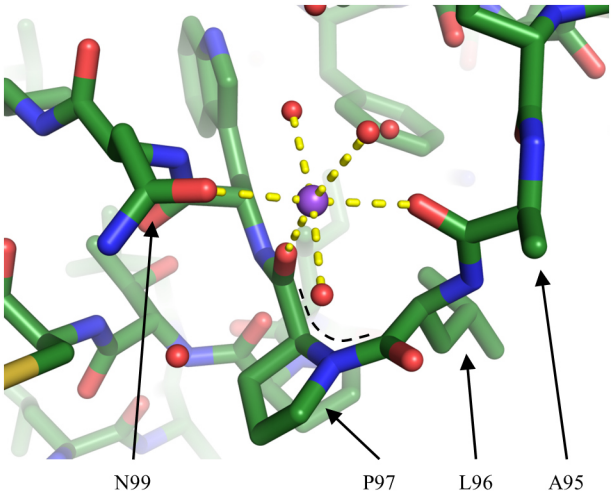


Figure 125: The cis-peptide bond (dashed black line) between residues L96 and P97 from PDB-entry 4NAW. The cis-configuration is stabilized by an unspecified ion, which contacts the backbone oxygens of A95 and P97, as well as the sidechain oxygen of N99. Three water molecules complete an octahedral coordination sphere.

A SAXS experiment was performed to determine the oligomerization state of ATG5 in solution (Figure 80). From the SAXS data, a MW of 39.8 kDa was determined, which corresponded to the MW of an ATG5 monomer of 33.1 kDa with a deviation of 20%. Fitting of the calculated scattering curves of an ATG5 monomer to the experimental scattering curves resulted in a good fit with a χ^2-value of 2.8 (Figure 80 c). At larger angles, corresponding to higher resolution, the calculated scattering curve didn't fit the experimental curve perfectly. The reason might have been, that the overall shape of ATG5 in solution and the ATG5 crystal structure was similar, but differed in details. The PISA server confirmed that ATG5 probably did not form oligomers. Taken these results together, ATG5 was probably monomeric in solution.

5.3.2 ATG5 phosphomimetic mutants structure

ATG5 residue T75 can be phosphorylated, resulting in the inhibition of the fusion of autophagosomes and lysosomes, thus inhibiting autophagy. ATG5 T75 was conserved across the kingdom of animals. In fungi and yeast it was usually replaced by hydrophobic amino acids like valine and isoleucine, only *S. cerevisiae* had a threonine in this position. In plants it was replaced by other amino acids, like phenylalanine in *A. thaliana* and leucine in rice. The structures of two ATG5 homologues were available in the PDB. In *S. cerevisiae* Atg5 T79 occupied the same space as T75 in the human form. Interestingly, in *K. marxianus*, also a yeast, Atg5 residue E78 replaced T75, which resembled the phosphomimetic mutant of the human form.

One possible effect of T75 phosphorylation was a changed affinity of ATG5 for ATG16L1 or TECPR-1, by influencing its conformation. To investigate the influence of T75 phosphorylation on the ATG5 structure, a phosphomimetic T75E mutant as well as a T75A negative control were generated. These two constructs were expressed, purified, crystallized and their structures were solved (Figure 81-Figure 85).

The structures of ATG5wt, T75E and T75A were extremely similar (Figure 126). RMSD-values for the backbone residues were 0.24 Å for T75E compared to wt and 0.22 Å for T75A compared to wt. For all atoms the values were 0.27 Å and 0.25 Å for T75E and T75A compared to wt respectively. Minor differences to the wild type structure were seen only in surface exposed loops and side chains, which had a higher degree of freedom. Like for the structure of ATG5wt, in T75E and T75A the first and the last three amino acids were not modelled, as well as an internal loop comprising residues 229-234, due to missing electron density, which indicated that these residues were flexible.

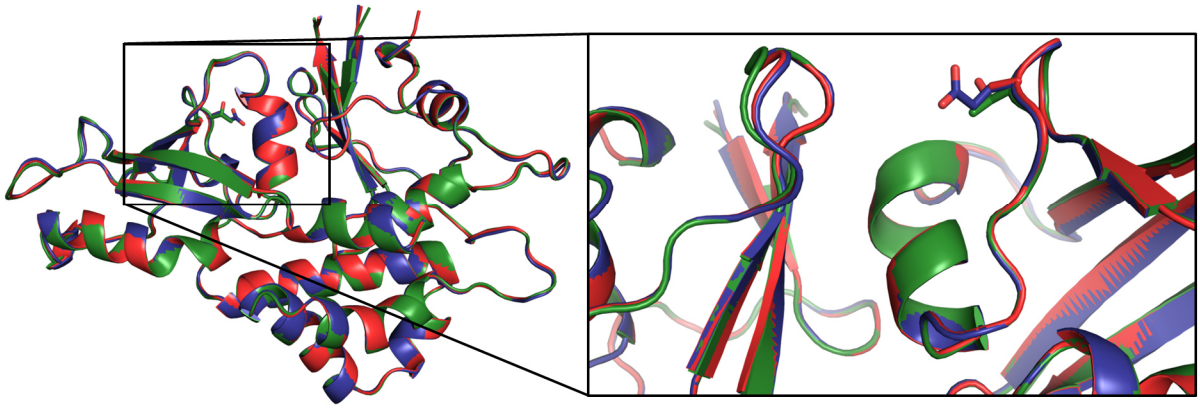


Figure 126: ATG5 structures: wt (green), T75E (blue), T75A (red). Minor differences were observed in the conformation of some surface loops. The inset shows the surrounding of residue 75, which is shown as a stick model for the three ATG5 variants. In comparison, the structures were extremely similar, also in proximity to the mutation.

The mutations were clearly visible in the electron density maps (Figure 127 a-c). If in ATG5T75E a threonine was modelled instead of the glutamate at position 75, additional unexplained electron density appeared around this residue (Figure 127 d). If a threonine was modelled instead of an alanine at position 75 in ATG5T75A, negative difference electron density appeared in proximity, indicating that no side chain atoms beyond C_β were present (Figure 127 e). This data confirmed the presence of the mutations.

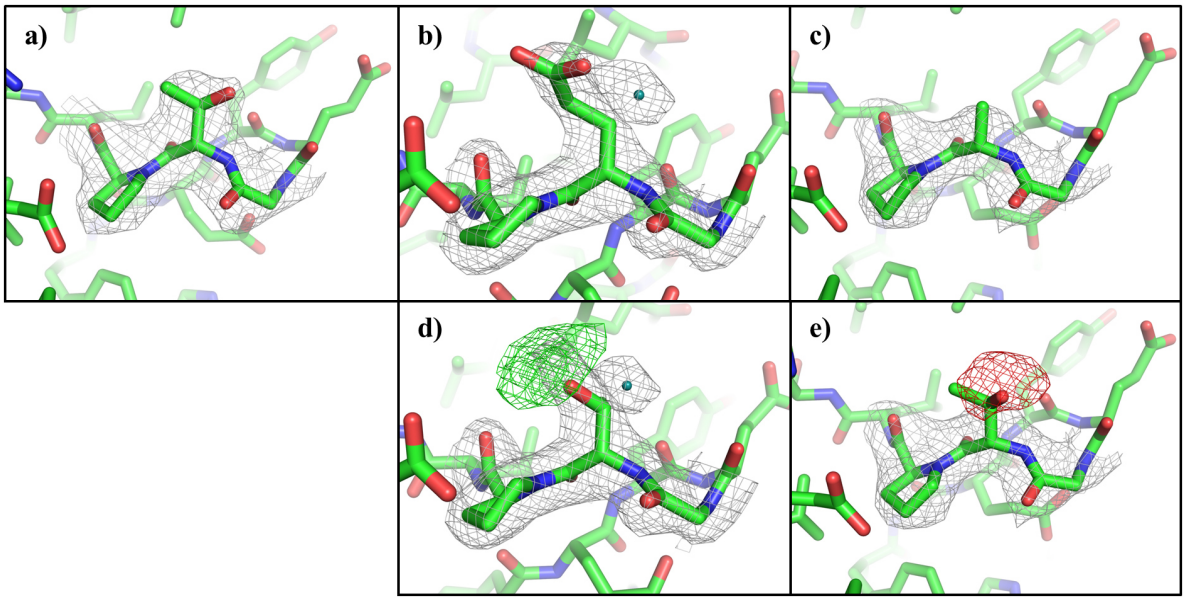


Figure 127: Electron density maps around residue 75. $2F_o$-F_c maps are shown in grey and are contoured at a level of 1.0 σ. F_o-F_c maps are shown in green and red and are contoured at a level of +3.0 and -3.0 σ respectively. a) ATG5wt, b) ATG5T75E, c) ATG5T75A, d) ATG5T75E, residue 75 is modelled as a threonine, e) ATG5T75A, residue 75 is modelled as a threonine. The results confirmed that the mutations were present in the respective structures.

In summary the different mutant variants of ATG5 were crystallized and the mutations were visible in their crystal structures. No influence of the phosphomimetic mutation T75E on the structure of ATG5 was observed, indicating that the phosphorylation of ATG5 probably wouldn't change its structure.

However, a direct influence of the phosphorylation of T75 on the binding of ATG16 was still possible, for instance by the destabilization of the tertiary structure of ATG5, which was not necessarily observable in the crystal structure. To address this question, the different ATG5 variants were examined for their biophysical properties by a thermofluor-assay (Figure 86) and limited proteolysis (Figure 87), but also in these experiments the three mutant ATG5 forms had identical properties. The melting temperatures of all ATG5 variants were in the range of 38-39°C and no differences in the degradation

patterns of the limited proteolysis experiment were observed. These results indicated that the T75 phosphorylation didn't destabilize ATG5.

5.3.3 Interaction ATG5 – ATG16L1

To test whether there was any influence of the T75 phosphorylation on the binding of ATG16, in the next steps binding studies of the different ATG5 variants with different ATG16 peptides, all comprising the ATG16 AFIM, were performed. In ITC experiments all ATG5 variants bound the ATG16L1(11-43) peptide with an affinity of about 1000 nM, the GST-ATG16L1(1-43) fusion protein with an affinity of about 300 nM and the GST-ATG16L1(1-30) fusion protein with an affinity of about 20 nM (Figure 105, Table 21).

Table 21: Binding affinities of the different ATG5 variants to the different ATG16L1 variants, used in ITC and BLI experiments. BLI experiments were run in triplets, so the indicated K_D-value are averages with standard deviations. No differences in the binding affinities of ATG16L1 to the different ATG5 variants were detected.

		ITC K_D [nM]	BLI K_D [nM]
GST-Atg16L1(1-30)	Atg5wt	10	50±5
	Atg5T75E	19	43±1
	Atg5T75A	21	42±7
GST-Atg16L1(1-43)	Atg5wt	250	
	Atg5T75E	320	
	Atg5T75A		
Atg16L1(11-43)	Atg5wt	1050	
	Atg5T75E	870	
	Atg5T75A	1100	

Interestingly the affinity of the GST-ATG16L1(1-30) fusion protein to ATG5 (10 nM magnitude) was higher than the affinity of GST-ATG16L1(1-43) fusion protein (100 nM magnitude) and ATG16L1(11-43) peptide (1000 nM magnitude). The structure of an N-terminal fragment of ATG16L1 in complex with ATG5 had previously been published (Figure 128). Most polar interactions were present between ATG5 and ATG16L1 residues 10-30. Apart from these, there were only two further contacts: ATG16L1 Y39 contacted the backbone oxygen of ATG5 P34 and ATG16L1 N40 contacted the side chain of ATG5 Y35. All residues of the ATG16L1 AFIM formed an α-helix, with a visible kink approximately at residue 30. The two contacts of the ATG16L1 residues 39 and 40 might have forced the helix in such a kinked conformation, thereby applying stress on the N-terminal part of the helix. This could have weakened the overall affinity of ATG16L1 to ATG5 and might have explained why the 1-30 construct had a higher affinity than the 1-43 construct. However, it did not explain the 10x lower affinity of ATG16L1(11-43) compared to GST-ATG16L1(1-43). Even though, individual GST did not interact with ATG5 in a control experiment, the presence of GST might have had an influence on the interaction of the fused ATG16L1 peptide with ATG5, by increasing the solubility of the peptide or by stabilizing the conformation of the α-helix. Also ATG16L1 residues 1-10 might have had an influence on the interaction with ATG5.

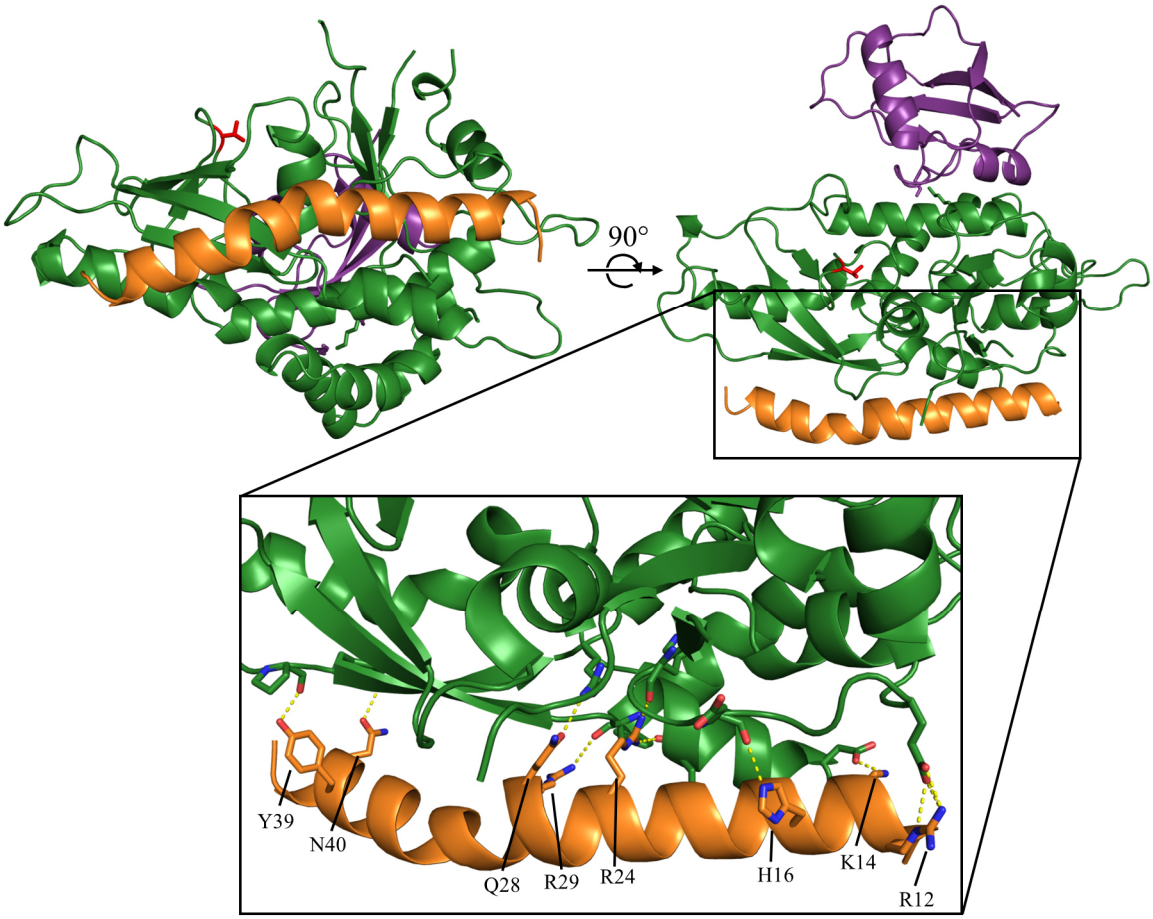


Figure 128: Human ATG5 (green), ATG12 (purple) and ATG16L1 (orange) from PDB-entry 4GDK. ATG5 residues T75 is highlighted in red. The inset shows the contacts between ATG5 and ATG16L1 (yellow dotted lines). The ATG16L1 residues, implicated in the interaction, are indicated below.

A part of the results from the ITC experiments were confirmed by BLI experiments, where all ATG5 variants bound the GST-ATG16L1(1-30) fusion protein with an affinity of about 50 nM (Table 21). This was lower than the about 20 nM measured by ITC, but still of the same magnitude. ITC is a method, where proteins interact free in solution, while for using BLI one of the binding partners has to be immobilized on a surface. Therefore diffusion is limited in BLI experiments and the resulting binding affinities can differ from the values obtained by other methods.

In conclusion all ATG5 variants showed similar binding affinities to the different ATG16L1 variants, used in these experiments. Hence the phosphorylation of T75 did not have a direct effect on the binding of ATG16L1. As can be seen in Figure 128, T75 and the binding site of ATG16L1 lie on the surface of ATG5, separated by an angle of about 90°. A direct influence of the phosphorylation of T75 on the binding of ATG16L1 would most probably have had to be transmitted by a conformational change, which was not observed in the crystal structures, or by long range charged interactions. With the data presented here, a direct influence of the T75 phosphorylation on the ATG16L1 interaction *in vitro* could be excluded. Therefore the phosphorylation of ATG5 T75 must have had a different, less direct effect on autophagy, probably by influencing the interaction of ATG5 with one of its many binding partners, not covered in this work.

6 Conclusions and Outlook

The main objective of this work, the characterization of the binding interfaces of VirG/IcsB and VirG/ATG5 for the identification of an inhibitor of the *Shigella* camouflage mechanism, has not been accomplished. The main problem was the difficult expression of the VirG protein, but also the late identification of the putative VirGE76-IcsB interaction and the missing interaction between VirG and ATG5. However, some important steps have been taken on the way to a better understanding of the *Shigella* camouflage mechanism.

Eight initial VirG-constructs were expressed as Strep- and GST-tagged proteins, but soluble protein expression was low. In further experiments these eight constructs were coexpressed with the published interaction partners IcsB and ATG5, as well as periplasmically expressed, without an improvement of the soluble protein yield. SEC showed that the fraction of solubly expressed protein formed aggregates and was therefore not suitable to set up crystallization experiments. For this reason the ESPRIT system was deployed to screen for soluble VirG-constructs, resulting in the identification of 18 interesting constructs that covered parts of VirG, where no structural information was available before, 13 of which covered parts of the IcsB/ATG5 binding domain. One of these constructs, VirGE76, comprising residues 340-758, was expressed solubly and was used in interaction studies with IcsB and ATG5, as it contained the whole IcsB/ATG5 binding domain, except for the first 20 aa. Another ESPRIT-construct, VirGE94, comprising residues 419-758, was crystallized and its structure was solved.

The VirGE94 structure contained the previously published structure of the VirG autochaperone domain, a continuation of the β-helix by about four turns and a previously unknown domain, here termed the VirG insertion domain (VID), sitting on top of the β-helix. The previously described polar localization region has been identified to be part of the β-helical backbone of VirG. It contained a negatively charged pocket that might be involved in the interaction with an unknown binding partner that is responsible for the unipolar localization of VirG. The VID interacted with the β-helical backbone *via* a hydrophobic interface. The fold of the VID was identified by the DALI server to be similar to the folds of the autochaperone domains of different autotransporter proteins. For this reason a potential function as a second autochaperone domain, nucleating the fold of a continuation of the β-helix, has been postulated. If the location and conformation of the VID in the full length VirG α-domain was similar to its position in the VirGE94 crystal structure, it would have introduced a 90° kink in the β-helical backbone. SAXS experiments to elucidate the position of the VID in the longer VirGE76 construct were not successful, due to aggregation of the protein. However, SAXS data confirmed that VirGE94 formed dimers in solution, which were mediated by the interaction of the VIDs. This dimerization of VirGE94 was most probably an artifact of this construct that does not have an *in vivo* function in the previously described oligomerization of VirG on the outer bacterial membrane. The reason for this assumption is that in the longer VirGE76 construct, dimerization was abolished, which might be explained by an autochaperone function of the VID, as the dimerization interface would be blocked by the following turns of the β-helix.

The autochaperone domain was identified to be essential for the stability of the VirG α-domain, as no soluble expression of constructs, lacking this domain, has been observed. On the other hand, N-terminal residues prior to 340 were found to destabilize the VirG α-domain, as the ESPRIT screening didn't yield constructs with N-termini in this region and also other constructs with N-termini in this section did not express solubly. Finally 35 new VirG-constructs were designed to elucidate the structure of further N-terminal parts of the VirG α-domain, but none of them crystallized.

Initially twelve different constructs of IcsB were expressed and purified, which mostly yielded high amounts of pure and soluble protein, but didn't crystallize. The application of different rescue strategies like surface entropy reduction, surface lysine methylation and *in situ* proteolysis didn't improve the

crystallizability. 19 new constructs were designed, based on experimentally determined boundaries from limited proteolysis, but didn't crystallize either. CD-spectroscopy results showed that IcsB contains a major fraction of random coil secondary structure elements, which might have hindered crystallization.

The IcsB-CBD has been mapped to residues 26-125 in pulldown experiments, which were confirmed by a bioinformatics prediction, SAXS experiments and by the protection of the binding site in a limited proteolysis experiment. The position of the CBD is in conflict with the previously published binding site between residues 171-247. However, due to the multitude of methods used in this work, the N-terminal position of the CBD is more probable.

The presence of the CBD destabilized the respective IcsB-constructs, if IpgA was not coexpressed. A GST-tagged IcsB-construct, covering the CBD, aggregated, even though the IcsB-construct was small compared to the solubility tag GST. On the other hand, IcsB-constructs, lacking the CBD, were expressed solubly in the absence of IpgA. These results were confirmed, by the formation of aggregates of IcsB, upon removal of IpgA by IEC, and by thermofluor-assays, where the IcsB-IpgA complex had an intermediate melting temperature of 50°C, between the melting temperatures of individual IcsB at 37°C and IpgA at 72°C. As IpgA is stripped off IcsB, upon secretion by the TTSS, the CBD might have an important *in vivo* function, which needs further characterization.

C-terminally truncated constructs of IcsB could not be expressed solubly, indicating the presence of a C-terminal domain, important for the stability of IcsB. The interaction of IcsB with the published interaction partners Toca-1 and cholesterol could not be confirmed, which might have been an effect of the limitations of the IcsB-constructs used in this work (complex with IpgA). The results of a Phyre2 prediction of the tertiary structure of IcsB have been used to propose an interaction of IcsB and the N-WASP-GBD, with an activatory effect on N-WASP. Indications for this interaction have been observed in a thermofluor-assay, using N-WASP–IcsB fusion proteins. However, because of the problems in this experiment and due to the fact that the results could not be confirmed in pulldown experiments, this interaction needs to be validated by an independent method.

Individual human ATG5 was expressed solubly, purified, crystallized and its structure was solved. It showed two ubiquitin-like fold domains and an α-helical bundle region, as observed in the structure of ATG5 from the human ATG5-12/16 complex. ATG5 was confirmed to be monomeric in solution by SAXS. The molecular basis of the function of T75 phosphorylation was investigated by solving the structures of a phosphomimetic mutant T75E as well as a control mutant T75A and comparing them to the structure of the wild type protein. Between the ATG5 variants, no differences in their structures, their stabilities or their binding to ATG16L1 were observed, indicating a less direct effect of the T75 phosphorylation on autophagy, e.g. *via* an unidentified interaction partner.

Interaction studies were performed to test the interactions between VirG, IcsB and ATG5, which formed the basis of the *Shigella* camouflage mechanism. Between IcsB and VirGE76 and interaction with a K_D of 600±205 nM has been observed by MST. This interaction could not be confirmed in pulldown studies, for this reason it still needs to be verified by an independent method. No interaction was observed between IcsB and VirGE94 as well as ATG5 and any VirG-construct. The reason for the missing interaction of ATG5 and VirG might have been the missing first 20 aa of the IcsB/ATG5 binding domain. It is also possible that other factors are involved in the *Shigella* camouflage mechanism, e.g. up to date unidentified binding partners.

One possibility might be the involvement of the septin cytoskeleton. In 2010 it was shown that septins form rings around the actin tails of *Shigella* and later also around the bacterial cells themselves, the so called septin cages. Those cages consist of alternate rings of actin and septins, with septin caps at the bacterial cell poles. *In vivo*, bacteria are either motile or confined by septin cages, hence these cages

might be a way to restrict the ABM and dissemination of *Shigella*. Enclosed bacteria recruit LC3 and p62, indicating that these cells are targeted for autophagic degradation [216]. These results indicate that the ABM itself might be a part of the evasion mechanism of *Shigella*, as it is difficult to enclose a quickly moving object in an autophagosome. The human cell first has to stop the bacterium, which is accomplished by the septin cages. These results also indicate that the actin tails, produced by intracellular *Shigella*, might represent a PAMP that is recognized by the innate immune response and leads to the targeting of *Shigella* cells to autophagic degradation. The interplay between the components of the septin cytoskeleton, the actin cytoskeleton, the ABPs and *Shigella* promise to be an interesting field of research in the future.

Also the regulation of actin polymerization by the Arp2/3 complex and by formins in ABM, the interplay between the two systems and the role of VirG and IcsB in these processes might be an interesting research target. Formins have been shown to have a function in the formation of protrusions. IpgB2 has been shown to activate the formin Dia1, but is not essential for the formation of protrusions. The human protein Toca-1, a known regulator of both the Arp2/3 complex and formins, is recruited to the surface of *Shigella* by IcsB, indicating that IcsB might be involved in the regulation of actin polymerization by formins. The putative activating effect of IcsB on N-WASP, which is the activator of the Arp2/3 complex, indicates that IcsB might have a role in the regulation of ABM by both the Arp2/3 and the formin system.

In the future, the VirG-constructs, generated in this work, espacially VirGE76, might be used for interaction studies, using soluble, purified proteins, to identify VirG interaction partners and regions. Should the interaction between VirGE76 and IcsB be verified, coexpression and cocrystallization experiments, using these two proteins could be performed. The soluble VirG constructs might also be cocrystallized with DOC, to identify the binding site and characterize possible conformational changes. Also crystallization of VirGK40 might be tested, which was not possible anymore in the time frame of this work. For the further characterization of VirG, the SAXS experiments using VirGE76, which failed due to aggregation of the protein, could be repeated, to determine the location of the VID and the conformation of the β-helical backbone. Also a new screening for soluble constructs, using the C-terminus of the VID as the C-terminal boundary, by the ESPRIT system or a similar method, might be performed. Soluble VirG-constructs are perfect targets for the development of vaccines [217], which is being tested at the moment in collaboration with Prof. Dr. Carlos Guzmán from the HZI in Braunschweig. For IcsB the putative interaction with the N-WASP-GBD needs to be verified by an independent interaction method.

Finally the results from this work will hopefully help to shed light on the molecular basis of the *Shigella* camouflage mechanism in the future.

References

[1] I. Kim, S. Rodriguez-Enriquez, and J. J. Lemasters, “Selective degradation of mitochondria by mitophagy,” *Arch. Biochem. Biophys.*, vol. 462, no. 2, pp. 245–253, 2007.

[2] Y. Sakai, M. Oku, I. J. van der Klei, and J. A. K. W. Kiel, “Pexophagy: autophagic degradation of peroxisomes,” *Biochim. Biophys. Acta (BBA)-Molecular Cell Res.*, vol. 1763, no. 12, pp. 1767–1775, 2006.

[3] G. Bjørkøy, T. Lamark, A. Brech, H. Outzen, M. Perander, A. Øvervatn, H. Stenmark, and T. Johansen, “p62/SQSTM1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death,” *J. Cell Biol.*, vol. 171, no. 4, pp. 603–614, 2005.

[4] V. Deretic, “Autophagy in immunity and cell-autonomous defense against intracellular microbes,” *Immunol. Rev.*, vol. 240, no. 1, pp. 92–104, 2011.

[5] A. B. Strawbridge and J. S. Blum, “Autophagy in MHC class II antigen processing,” *Curr. Opin. Immunol.*, vol. 19, no. 1, pp. 87–92, 2007.

[6] D. Glick, S. Barth, and K. F. Macleod, “Autophagy: cellular and molecular mechanisms,” *J. Pathol.*, vol. 221, no. 1, pp. 3–12, 2010.

[7] M. Hamasaki, N. Furuta, A. Matsuda, A. Nezu, A. Yamamoto, N. Fujita, H. Oomori, T. Noda, T. Haraguchi, Y. Hiraoka, and others, “Autophagosomes form at ER-mitochondria contact sites,” *Nature*, vol. 495, no. 7441, pp. 389–393, 2013.

[8] A. Kuma and N. Mizushima, “Physiological role of autophagy as an intracellular recycling system: with an emphasis on nutrient metabolism,” in *Seminars in cell & developmental biology*, 2010, vol. 21, no. 7, pp. 683–690.

[9] V. Deretic and B. Levine, “Autophagy, immunity, and microbial adaptations,” *Cell Host Microbe*, vol. 5, no. 6, pp. 527–549, 2009.

[10] D. J. Klionsky, J. M. Cregg, W. A. Dunn, S. D. Emr, Y. Sakai, I. V Sandoval, A. Sibirny, S. Subramani, M. Thumm, M. Veenhuis, and others, “A unified nomenclature for yeast autophagy-related genes,” *Dev. Cell*, vol. 5, no. 4, pp. 539–545, 2003.

[11] F. Reggiori and D. J. Klionsky, “Autophagic processes in yeast: mechanism, machinery and regulation,” *Genetics*, vol. 194, no. 2, pp. 341–361, 2013.

[12] A. Kihara, T. Noda, N. Ishihara, and Y. Ohsumi, “Two Distinct Vps34 Phosphatidylinositol 3--Kinase Complexes Function in Autophagy and Carboxypeptidase Y Sorting inSaccharomyces cerevisiae,” *J. Cell Biol.*, vol. 152, no. 3, pp. 519–530, 2001.

[13] Y. Araki, W.-C. Ku, M. Akioka, A. I. May, Y. Hayashi, F. Arisaka, Y. Ishihama, and Y. Ohsumi, “Atg38 is required for autophagy-specific phosphatidylinositol 3-kinase complex integrity,” *J. Cell Biol.*, vol. 203, no. 2, pp. 299–313, 2013.

[14] Z. Xie and D. J. Klionsky, “Autophagosome formation: core machinery and adaptations,” *Nat. Cell Biol.*, vol. 9, no. 10, pp. 1102–1109, 2007.

[15] N. Mizushima, A. Yamamoto, M. Hatano, Y. Kobayashi, Y. Kabeya, K. Suzuki, T. Tokuhisa, Y. Ohsumi, and T. Yoshimori, “Dissection of autophagosome formation using Apg5-deficient mouse embryonic stem cells,” *J. Cell Biol.*, vol. 152, no. 4, pp. 657–668, 2001.

[16] N. Mizushima, T. Noda, T. Yoshimori, Y. Tanaka, T. Ishii, M. D. George, D. J. Klionsky, M. Ohsumi, and Y. Ohsumi, “A protein conjugation system essential for autophagy,” *Nature*, vol. 395, no. 6700, pp. 395–398, 1998.

[17] I. Tanida, N. Mizushima, M. Kiyooka, M. Ohsumi, T. Ueno, Y. Ohsumi, and E. Kominami, “Apg7p/Cvt2p: a novel protein-activating enzyme essential for autophagy,” *Mol. Biol. Cell*, vol. 10, no. 5, pp. 1367–1379, 1999.

[18] T. Shintani, N. Mizushima, Y. Ogawa, A. Matsuura, T. Noda, and Y. Ohsumi, “Apg10p, a novel protein-conjugating enzyme essential for autophagy in yeast,” *EMBO J.*, vol. 18, no. 19, pp. 5234–5241, 1999.

[19] Y. Fujioka, N. N. Noda, H. Nakatogawa, Y. Ohsumi, and F. Inagaki, “Dimeric coiled-coil structure of Saccharomyces cerevisiae Atg16 and its functional significance in autophagy,” *J. Biol. Chem.*, vol. 285, no. 2, pp. 1508–1515, 2010.

[20] A. M. Schalk, “Structural and functional characterization of the autophagy proteins Atg5 and Atg16L1 and their interaction partners,” 2011.

[21] Y. Ichimura, T. Kirisako, T. Takao, Y. Satomi, Y. Shimonishi, N. Ishihara, N. Mizushima, I. Tanida, E. Kominami, M. Ohsumi, and others, “A ubiquitin-like system mediates protein lipidation,” *Nature*, vol. 408, no. 6811, pp. 488–492, 2000.

[22] T. Hanada, N. N. Noda, Y. Satomi, Y. Ichimura, Y. Fujioka, T. Takao, F. Inagaki, and Y. Ohsumi, “The Atg12-Atg5 conjugate has a novel E3-like activity for protein lipidation in autophagy,” *J. Biol. Chem.*, vol. 282, no. 52, pp. 37298–37302, 2007.

[23] N. Mizushima, H. Sugita, T. Yoshimori, and Y. Ohsumi, “A new protein conjugation system in human The counterpart of the yeast Apg12p conjugation system essential for autophagy,” *J. Biol. Chem.*, vol. 273, no. 51, pp. 33889–33892, 1998.

[24] K. Ishibashi, N. Fujita, E. Kanno, H. Omori, T. Yoshimori, T. Itoh, and M. Fukuda, “Atg16L2, a novel isoform of mammalian Atg16L that is not essential for canonical autophagy despite forming an Atg12--5-16L2 complex,” *Autophagy*, vol. 7, no. 12, pp. 1500–1513, 2011.

[25] T. Shpilka, H. Weidberg, S. Pietrokovski, and Z. Elazar, “Atg8: an autophagy-related ubiquitin-like protein family,” *Genome Biol*, vol. 12, no. 7, p. 226, 2011.

[26] J. Geng and D. J. Klionsky, “The Atg8 and Atg12 ubiquitin-like conjugation systems in macroautophagy,” *EMBO Rep.*, vol. 9, no. 9, pp. 859–864, 2008.

[27] A. Kuma, N. Mizushima, N. Ishihara, and Y. Ohsumi, “Formation of the~ 350-kDa Apg12-Apg5{\textperiodcentered} Apg16 multimeric complex, mediated by Apg16 oligomerization, is essential for autophagy in yeast,” *J. Biol. Chem.*, vol. 277, no. 21, pp. 18619–18625, 2002.

[28] N. Fujita, T. Itoh, H. Omori, M. Fukuda, T. Noda, and T. Yoshimori, “The Atg16L complex specifies the site of LC3 lipidation for membrane biogenesis in autophagy,” *Mol. Biol. Cell*, vol. 19, no. 5, pp. 2092–2100, 2008.

[29] N. Mizushima, A. Kuma, Y. Kobayashi, A. Yamamoto, M. Matsubae, T. Takao, T. Natsume, Y. Ohsumi, and T. Yoshimori, “Mouse Apg16L, a novel WD-repeat protein, targets to the autophagic isolation membrane with the Apg12-Apg5 conjugate,” *J. Cell Sci.*, vol. 116, no. 9, pp. 1679–1688, 2003.

[30] H. Nakatogawa, Y. Ichimura, and Y. Ohsumi, “Atg8, a ubiquitin-like protein required for autophagosome formation, mediates membrane tethering and hemifusion,” *Cell*, vol. 130, no. 1, pp. 165–178, 2007.

[31] Z. Xie, U. Nair, and D. J. Klionsky, “Atg8 controls phagophore expansion during autophagosome formation,” *Mol. Biol. Cell*, vol. 19, no. 8, pp. 3290–3298, 2008.

[32] Y. Kabeya, N. Mizushima, T. Ueno, A. Yamamoto, T. Kirisako, T. Noda, E. Kominami, Y. Ohsumi, and T. Yoshimori, “LC3, a mammalian homologue of yeast Apg8p, is localized in autophagosome membranes after processing,” *EMBO J.*, vol. 19, no. 21, pp. 5720–5728, 2000.

[33] Y. I. Xu, C. Jagannath, X.-D. Liu, A. Sharafkhaneh, K. E. Kolodziejska, and N. T. Eissa, “Toll-like receptor 4 is a sensor for autophagy associated with innate immunity,” *Immunity*, vol. 27, no. 1, pp. 135–144, 2007.

[34] M. A. Delgado, R. A. Elmaoued, A. S. Davis, G. Kyei, and V. Deretic, “Toll-like receptors control autophagy,” *EMBO J.*, vol. 27, no. 7, pp. 1110–1121, 2008.

[35] L. H. Travassos, L. A. M. Carneiro, M. Ramjeet, S. Hussey, Y.-G. Kim, J. G. Magalhães, L. Yuan, F. Soares, E. Chea, L. Le Bourhis, and others, “Nod1 and Nod2 direct autophagy by recruiting ATG16L1 to the plasma membrane at the site of bacterial entry,” *Nat. Immunol.*, vol. 11, no. 1, pp. 55–62, 2010.

[36] J.-O. Pyo, M.-H. Jang, Y.-K. Kwon, H.-J. Lee, J.-I. Jun, H.-N. Woo, D.-H. Cho, B. Choi, H. Lee, J.-H. Kim, and others, “Essential roles of Atg5 and FADD in autophagic cell death dissection of autophagic cell death into vacuole formation and cell death,” *J. Biol. Chem.*, vol. 280, no. 21, pp. 20722–20729, 2005.

[37] A. J. Perrin, X. Jiang, C. L. Birmingham, N. S. Y. So, and J. H. Brumell, “Recognition of bacteria in the cytosol of mammalian cells by the ubiquitin system,” *Curr. Biol.*, vol. 14, no. 9, pp. 806–811, 2004.

[38] P. S. Manzanillo, J. S. Ayres, R. O. Watson, A. C. Collins, G. Souza, C. S. Rae, D. S. Schneider, K. Nakamura, M. U. Shiloh, and J. S. Cox, “The ubiquitin ligase parkin mediates resistance to intracellular pathogens,” *Nature*, vol. 501, no. 7468, pp. 512–516, 2013.

[39] A. Huett, R. J. Heath, J. Begun, S. O. Sassi, L. A. Baxt, J. M. Vyas, M. B. Goldberg, and R. J. Xavier, “The LRR and RING domain protein LRSAM1 is an E3 ligase crucial for ubiquitin-dependent autophagy of intracellular Salmonella Typhimurium,” *Cell Host Microbe*, vol. 12, no. 6, pp. 778–790, 2012.

[40] S. Pankiv, T. H. Clausen, T. Lamark, A. Brech, J.-A. Bruun, H. Outzen, A. Øvervatn, G. Bjørkøy, and T. Johansen, “p62/SQSTM1 binds directly to Atg8/LC3 to facilitate degradation of ubiquitinated protein aggregates by autophagy,” *J. Biol. Chem.*, vol. 282, no. 33, pp. 24131–24145, 2007.

[41] T. L. M. Thurston, M. P. Wandel, N. von Muhlinen, Á. Foeglein, and F. Randow, “Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion,” *Nature*, vol. 482, no. 7385, pp. 414–418, 2012.

[42] L. A. Baxt, A. C. Garza-Mayers, and M. B. Goldberg, “Bacterial subversion of host innate immune pathways,” *Science (80-.).*, vol. 340, no. 6133, pp. 697–701, 2013.

[43] F. S. Mesquita, M. Thomas, M. Sachse, A. J. M. Santos, R. Figueira, and D. W. Holden, “The Salmonella deubiquitinase SseL inhibits selective autophagy of cytosolic aggregates,” *PLoS Pathog.*, vol. 8, no. 6, p. e1002743, 2012.

[44] S. S. Ivanov and C. R. Roy, “Modulation of ubiquitin dynamics and suppression of DALIS formation by the Legionella pneumophila Dot/Icm system,” *Cell. Microbiol.*, vol. 11, no. 2, pp. 261–278, 2009.

[45] A. Choy, J. Dancourt, B. Mugo, T. J. O’Connor, R. R. Isberg, T. J. Melia, and C. R. Roy, “The Legionella effector RavZ inhibits host autophagy through irreversible Atg8 deconjugation,” *Science (80-.).*, vol. 338, no. 6110, pp. 1072–1076, 2012.

[46] M. Ogawa, T. Yoshimori, T. Suzuki, H. Sagara, N. Mizushima, and C. Sasakawa, “Escape of intracellular Shigella from autophagy,” *Science (80-.).*, vol. 307, no. 5710, p. 727, 2005.

[47] L. Dortet, S. Mostowy, A. S. Louaka, E. Gouin, M.-A. Nahori, E. A. C. Wiemer, O. Dussurget, and P. Cossart, “Recruitment of the major vault protein by InlK: a Listeria monocytogenes strategy to avoid autophagy,” *PLoS Pathog.*, vol. 7, no. 8, p. e1002168, 2011.

[48] H. F. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, J. Darnell, and others, *Molecular cell biology*, vol. 4. WH Freeman New York, 2000.

[49] W. Kabsch, H. G. Mannherz, D. Suck, E. F. Pai, and K. C. Holmes, “Atomic structure of the actin: DNase I complex.,” *Nature*, no. 347, pp. 37–44, 1990.

[50] R. Dominguez and K. C. Holmes, “Actin structure and function,” *Annu. Rev. Biophys.*, vol. 40, p. 169, 2011.

[51] R. Dominguez, “Actin-binding proteins--a unifying hypothesis,” *Trends Biochem. Sci.*, vol. 29, no. 11, pp. 572–578, 2004.

[52] P. Graceffa and R. Dominguez, “Crystal Structure of Monomeric Actin in the ATP State STRUCTURAL BASIS OF NUCLEOTIDE-DEPENDENT ACTIN DYNAMICS,” *J. Biol. Chem.*, vol. 278, no. 36, pp. 34172–34180, 2003.

[53] T. Oda, M. Iwasa, T. Aihara, Y. Maéda, and A. Narita, “The nature of the globular-to fibrous-actin transition,” *Nature*, vol. 457, no. 7228, pp. 441–445, 2009.

[54] K. C. Holmes, D. Popp, W. Gebhard, W. Kabsch, and others, “Atomic model of the actin filament,” *Nature*, vol. 347, no. 6288, pp. 44–49, 1990.

[55] T. Fujii, A. H. Iwane, T. Yanagida, and K. Namba, “Direct visualization of secondary structures of F-actin by electron cryomicroscopy,” *Nature*, vol. 467, no. 7316, pp. 724–728, 2010.

[56] L. Blanchoin and T. D. Pollard, “Hydrolysis of ATP by polymerized actin depends on the bound divalent cation but not profilin,” *Biochemistry*, vol. 41, no. 2, pp. 597–602, 2002.

[57] R. Melki, S. Fievez, and M.-F. Carlier, “Continuous monitoring of Pi release following nucleotide hydrolysis in actin or tubulin assembly using 2-amino-6-mercapto-7-methylpurine ribonucleoside and purine-nucleoside phosphorylase as an enzyme-linked assay,” *Biochemistry*, vol. 35, no. 37, pp. 12038–12045, 1996.

[58] T. D. Pollard, “Rate constants for the reactions of ATP-and ADP-actin with the ends of actin filaments.,” *J. Cell Biol.*, vol. 103, no. 6, pp. 2747–2754, 1986.

[59] A. Wegner and G. Isenberg, “12-fold difference between the critical monomer concentrations of the two ends of actin filaments in physiological salt conditions,” *Proc. Natl. Acad. Sci.*, vol. 80, no. 16, pp. 4922–4925, 1983.

[60] T. D. Pollard and G. G. Borisy, “Cellular motility driven by assembly and disassembly of actin filaments,” *Cell*, vol. 112, no. 4, pp. 453–465, 2003.

[61] T. D. Pollard, L. Blanchoin, and R. D. Mullins, “Molecular mechanisms controlling actin filament dynamics in nonmuscle cells,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 29, no. 1, pp. 545–576, 2000.

[62] T. D. Pollard, “Regulation of actin filament assembly by Arp2/3 complex and formins,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 36, pp. 451–477, 2007.

[63] R. C. Robinson, K. Turbedsky, D. A. Kaiser, J.-B. Marchand, H. N. Higgs, S. Choe, and T. D. Pollard, “Crystal structure of Arp2/3 complex,” *Science (80-.).*, vol. 294, no. 5547, pp. 1679–1684, 2001.

[64] I. Rouiller, X.-P. Xu, K. J. Amann, C. Egile, S. Nickell, D. Nicastro, R. Li, T. D. Pollard, N. Volkmann, and D. Hanein, “The structural basis of actin filament branching by the Arp2/3 complex,” *J. Cell Biol.*, vol. 180, no. 5, pp. 887–895, 2008.

[65] A. Schönichen and M. Geyer, “Fifteen formins for an actin filament: a molecular view on the regulation of human formins,” *Biochim. Biophys. Acta (BBA)-Molecular Cell Res.*, vol. 1803, no. 2, pp. 152–163, 2010.

[66] M. Lammers, R. Rose, A. Scrima, and A. Wittinghofer, “The regulation of mDia1 by autoinhibition and its release by Rho• GTP,” *EMBO J.*, vol. 24, no. 23, pp. 4176–4187, 2005.

[67] T. D. Pollard, W. C. Earnshaw, and J. Lippincott-Schwartz, *Cell biology*. Elsevier Health Sciences, 2007.

[68] M. E. Quinlan, J. E. Heuser, E. Kerkhoff, and R. D. Mullins, “Drosophila Spire is an actin nucleation factor,” *Nature*, vol. 433, no. 7024, pp. 382–388, 2005.

[69] H. Miki, S. Suetsugu, and T. Takenawa, “WAVE, a novel WASP-family protein involved in actin reorganization induced by Rac,” *EMBO J.*, vol. 17, no. 23, pp. 6932–6941, 1998.

[70] H. Miki, K. Miura, and T. Takenawa, “N-WASP, a novel actin-depolymerizing protein, regulates the cortical cytoskeletal rearrangement in a PIP2-dependent manner downstream of tyrosine kinases.,” *EMBO J.*, vol. 15, no. 19, p. 5326, 1996.

[71] R. Rohatgi, L. Ma, H. Miki, M. Lopez, T. Kirchhausen, T. Takenawa, and M. W. Kirschner, “The interaction between N-WASP and the Arp2/3 complex links Cdc42-dependent signals to actin assembly,” *Cell*, vol. 97, no. 2, pp. 221–231, 1999.

[72] E. Derivery and A. Gautreau, “Generation of branched actin networks: assembly and regulation of the N-WASP and WAVE molecular machines,” *Bioessays*, vol. 32, no. 2, pp. 119–131, 2010.

[73] A. S. Kim, L. T. Kakalis, N. Abdul-Manan, G. A. Liu, and M. K. Rosen, “Autoinhibition and activation mechanisms of the Wiskott--Aldrich syndrome protein,” *Nature*, vol. 404, no. 6774, pp. 151–158, 2000.

[74] H.-Y. H. Ho, R. Rohatgi, A. M. Lebensohn, J. Li, S. P. Gygi, M. W. Kirschner, and others, “Toca-1 mediates Cdc42-dependent actin nucleation by activating the N-WASP-WIP complex,” *Cell*, vol. 118, no. 2, pp. 203–216, 2004.

[75] A. Frost, V. M. Unger, and P. De Camilli, “The BAR domain superfamily: membrane-molding macromolecules,” *Cell*, vol. 137, no. 2, pp. 191–196, 2009.

[76] R. Fricke, C. Gohl, E. Dharmalingam, A. Grevelhörster, B. Zahedi, N. Harden, M. Kessels, B. Qualmann, and S. Bogdan, “Drosophila Cip4/Toca-1 integrates membrane trafficking and actin dynamics through WASP and SCAR/WAVE,” *Curr. Biol.*, vol. 19, no. 17, pp. 1429–1437, 2009.

[77] P. Aspenström, N. Richnau, and A.-S. Johansson, “The diaphanous-related formin DAAM1 collaborates with the Rho GTPases RhoA and Cdc42, CIP4 and Src in regulating cell morphogenesis and actin dynamics,” *Exp. Cell Res.*, vol. 312, no. 12, pp. 2180–2194, 2006.

[78] P. Aspenström, “Formin-binding proteins: modulators of formin-dependent actin polymerization,” *Biochim. Biophys. Acta (BBA)-Molecular Cell Res.*, vol. 1803, no. 2, pp. 174–182, 2010.

[79] L. G. Tilney and D. A. Portnoy, “Actin filaments and the growth, movement, and spread of the intracellular bacterial parasite, Listeria monocytogenes.,” *J. Cell Biol.*, vol. 109, no. 4, pp. 1597–1608, 1989.

[80] C. Kocks, E. Gouin, M. Tabouret, P. Berche, H. Ohayon, and P. Cossart, “L. monocytogenes-induced actin assembly requires the actA gene product, a surface protein,” *Cell*, vol. 68, no. 3, pp. 521–531, 1992.

[81] M. D. Welch, J. Rosenblatt, J. Skoble, D. A. Portnoy, and T. J. Mitchison, “Interaction of human Arp2/3 complex and the Listeria monocytogenes ActA protein in actin filament nucleation,” *Science (80-.).*, vol. 281, no. 5373, pp. 105–108, 1998.

[82] N. Teysseire, C. Chiche-Portiche, and D. Raoult, “Intracellular movements of Rickettsia conorii adn R. typhi based on actin polymerization,” *Res. Microbiol.*, vol. 143, no. 9, pp. 821–829, 1992.

[83] R. A. Heinzen, S. F. Hayes, M. G. Peacock, and T. Hackstadt, “Directional actin polymerization associated with spotted fever group Rickettsia infection of Vero cells.,” *Infect. Immun.*, vol. 61, no. 5, pp. 1926–1935, 1993.

[84] S. Cudmore, P. Cossart, G. Griffiths, and M. Way, “Actin-based motility of vaccinia virus,” 1995.

[85] V. Moreau, F. Frischknecht, I. Reckmann, R. Vincentelli, G. Rabut, D. Stewart, and M. Way, “A complex of N-WASP and WIP integrates signalling cascades that lead to actin polymerization,” *Nat. Cell Biol.*, vol. 2, no. 7, pp. 441–448, 2000.

[86] 小河秀正, 中村明子, and 中谷林太郎, “Cinemicrographic study of tissue cell cultures infected with Shigella flexneri,” *Japanese J. Med. Sci. Biol.*, vol. 21, no. 4, pp. 259–273, 1968.

[87] S. Makino, C. Sasakawa, K. Kamata, T. Kurata, and M. Yoshikawa, “A genetic determinant required for continuous reinfection of adjacent cells on large plasmid in S. flexneri 2a,” *Cell*, vol. 46, no. 4, pp. 551–555, 1986.

[88] M. L. Bernardini, J. Mounier, H. D’Hauteville, M. Coquis-Rondon, and P. J. Sansonetti, “Identification of icsA, a plasmid locus of Shigella flexneri that governs bacterial intra-and intercellular spread through interaction with F-actin,” *Proc. Natl. Acad. Sci.*, vol. 86, no. 10, p. 3867, 1989.

[89] C. Egile, T. P. Loisel, V. Laurent, R. Li, D. Pantaloni, P. J. Sansonetti, and M.-F. Carlier, “Activation of the CDC42 effector N-WASP by the Shigella flexneri IcsA protein promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility,” *J. Cell Biol.*, vol. 146, no. 6, pp. 1319–1332, 1999.

[90] J. E. Heindl, I. Saran, C. Yi, C. F. Lesser, and M. B. Goldberg, “Requirement for formin-induced actin polymerization during spread of Shigella flexneri,” *Infect. Immun.*, vol. 78, no. 1, pp. 193–203, 2010.

[91] Robert Koch Institut, “SurvStat@RKI 2.0.” [Online]. Available: https://survstat.rki.de/Default.aspx. [Accessed: 26-Mar-2015].

[92] Center for Disease Control and Prevention, “National Shigella Surveillance.” [Online]. Available: http://www.cdc.gov/nationalsurveillance/shigella-surveillance.html. [Accessed: 26-Mar-2015].

[93] M. L. Bennish and B. J. Wojtyniak, “Mortality due to shigellosis: community and hospital data,” *Rev. Infect. Dis.*, vol. 13, no. Supplement 4, pp. S245–251, 1991.

[94] K. L. Kotloff, J. P. Winickoff, B. Ivanoff, J. D. Clemens, D. L. Swerdlow, P. J. Sansonetti, G. K. Adak, and M. M. Levine, “Global burden of Shigella infections: implications for vaccine development and implementation of control strategies,” *Bull. World Health Organ.*, vol. 77, no. 8, pp. 651–666, 1999.

[95] A. V Jennison and N. K. Verma, “Shigella flexneri infection: pathogenesis and vaccine development,” *FEMS Microbiol. Rev.*, vol. 28, no. 1, pp. 43–58, 2004.

[96] E. M. Barry, M. F. Pasetti, M. B. Sztein, A. Fasano, K. L. Kotloff, and M. M. Levine, “Progress and pitfalls in Shigella vaccine research,” *Nat. Rev. Gastroenterol. Hepatol.*, vol. 10, no. 4, pp. 245–255, 2013.

[97] R. S. Coimbra, F. Grimont, and P. A. D. Grimont, “Identification of Shigella serotypes by restriction of amplified O-antigen gene cluster,” *Res. Microbiol.*, vol. 150, no. 8, pp. 543–553, 1999.

[98] M. J. C. van den Beld and F. A. G. Reubsaet, “Differentiation between Shigella, enteroinvasive Escherichia coli (EIEC) and noninvasive Escherichia coli,” *Eur. J. Clin. Microbiol. Infect. Dis.*, pp. 1–6, 2011.

[99] S. Ashkenazi, I. Levy, V. Kazaronovski, and Z. Samra, “Growing antimicrobial resistance of Shigella isolates,” *J. Antimicrob. Chemother.*, vol. 51, no. 2, pp. 427–429, 2003.

[100] M. M. Venkatesan, M. B. Goldberg, D. J. Rose, E. J. Grotbeck, V. Burland, and F. R. Blattner, “Complete DNA sequence and analysis of the large virulence plasmid of Shigella flexneri,” *Infect. Immun.*, vol. 69, no. 5, p. 3271, 2001.

[101] C. Parsot, “Shigella type III secretion effectors: how, where, when, for what purposes?,” *Curr. Opin. Microbiol.*, vol. 12, no. 1, pp. 110–116, 2009.

[102] L. Zhu, G. Zhao, R. Stein, X. Zheng, W. Hu, N. Shang, X. Bu, X. Liu, J. Wang, E. Feng, and others, “The Proteome of Shigella flexneri 2a 2457T Grown at 30 and 37° C,” *Mol. Cell. Proteomics*, vol. 9, no. 6, p. 1209, 2010.

[103] L. Zhu, X. Liu, X. Zheng, X. Bu, G. Zhao, C. Xie, J. Zhang, N. Li, E. Feng, J. Wang, and others, “Global Analysis of a Plasmid-Cured Shigella flexneri Strain: New Insights into the Interaction between the Chromosome and a Virulence Plasmid,” *J. Proteome Res.*, vol. 9, no. 2, pp. 843–854, 2009.

[104] A. Blocker, P. Gounon, E. Larquet, K. Niebuhr, V. Cabiaux, C. Parsot, and P. Sansonetti, “The tripartite type III secreton of Shigella flexneri inserts IpaB and IpaC into host membranes,” *J. Cell Biol.*, vol. 147, no. 3, pp. 683–693, 1999.

[105] A. Blocker, N. Jouihri, E. Larquet, P. Gounon, F. Ebel, C. Parsot, P. Sansonetti, and A. Allaoui, “Structure and composition of the Shigella flexneri ‘needle complex’, a part of its type III secreton,” *Mol. Microbiol.*, vol. 39, no. 3, pp. 652–663, 2001.

[106] H. L. DuPont, M. M. Levine, R. B. Hornick, and S. B. Formal, “Inoculum size in shigellosis and implications for expected mode of transmission,” *J. Infect. Dis.*, pp. 1126–1128, 1989.

[107] N. E. Dickenson, L. Zhang, C. R. Epler, P. R. Adam, W. L. Picking, and W. D. Picking, “Conformational changes in IpaD from Shigella flexneri upon binding bile salts provide insight into the second step of type III secretion,” *Biochemistry*, vol. 50, no. 2, pp. 172–180, 2010.

[108] A. Brotcke Zumsteg, C. Goosmann, V. Brinkmann, R. Morona, and A. Zychlinsky, “IcsA Is a Shigella flexneri Adhesion Regulated by the Type III Secretion System and Required for Pathogenesis,” *Cell Host Microbe*, vol. 15, no. 4, pp. 435–445, 2014.

[109] J. S. Wassef, D. F. Keren, and J. L. Mailloux, “Role of M cells in initial antigen uptake and in ulcer formation in the rabbit intestinal loop model of shigellosis.,” *Infect. Immun.*, vol. 57, no. 3, pp. 858–863, 1989.

[110] A. Zychlinsky, M. C. Prevost, and P. J. Sansonetti, “Shigella flexneri induces apoptosis in infected macrophages,” *Nature*, vol. 358, no. 6382, pp. 167–169, 1992.

[111] A. Zychlinsky, B. Kenny, R. Ménard, M.-C. Prévost, I. B. Holland, and P. J. Sansonetti, “IpaB mediates macrophage apoptosis induced by Shigella flexneri,” *Mol. Microbiol.*, vol. 11, no. 4, pp. 619–627, 1994.

[112] P. Cossart and P. J. Sansonetti, “Bacterial invasion: the paradigms of enteroinvasive pathogens,” *Science (80-.).*, vol. 304, no. 5668, pp. 242–248, 2004.

[113] H. d’Hauteville, S. Khan, D. J. Maskell, A. Kussak, A. Weintraub, J. Mathison, R. J. Ulevitch, N. Wuscher, C. Parsot, and P. J. Sansonetti, “Two msbB genes encoding maximal acylation of lipid A are required for invasive Shigella flexneri to mediate inflammatory rupture and destruction of the intestinal epithelium,” *J. Immunol.*, vol. 168, no. 10, pp. 5240–5251, 2002.

[114] D. W. Kim, G. Lenzen, A.-L. Page, P. Legrain, P. J. Sansonetti, and C. Parsot, “The Shigella flexneri effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 102, no. 39, pp. 14046–14051, 2005.

[115] L. Arbibe, D. W. Kim, E. Batsche, T. Pedron, B. Mateescu, C. Muchardt, C. Parsot, and P. J. Sansonetti, “An injected bacterial effector targets chromatin access for transcription factor NF-κB to alter transcription of host genes involved in immune responses,” *Nat. Immunol.*, vol. 8, no. 1, pp. 47–56, 2007.

[116] J. Okuda, T. Toyotome, N. Kataoka, M. Ohno, H. Abe, Y. Shimura, A. Seyedarabi, R. Pickersgill, and C. Sasakawa, “Shigella effector IpaH9. 8 binds to a splicing factor U2AF 35 to modulate host immune responses,” *Biochem. Biophys. Res. Commun.*, vol. 333, no. 2, pp. 531–539, 2005.

[117] J. R. Rohde, A. Breitkreutz, A. Chenal, P. J. Sansonetti, and C. Parsot, “Type III secretion effectors of the IpaH family are E3 ubiquitin ligases,” *Cell Host Microbe*, vol. 1, no. 1, pp. 77–83, 2007.

[118] M. Ogawa, Y. Handa, H. Ashida, M. Suzuki, and C. Sasakawa, “The versatility of Shigella effectors,” *Nat. Rev. Microbiol.*, vol. 6, no. 1, pp. 11–16, 2008.

[119] N. Ramarao, C. Le Clainche, T. Izard, R. Bourdet-Sicard, E. Ageron, P. J. Sansonetti, M.-F. Carlier, and G. T. Van Nhieu, “Capping of actin filaments by vinculin activated by the Shigella IpaA carboxyl-terminal domain,” *FEBS Lett.*, vol. 581, no. 5, pp. 853–857, 2007.

[120] K. A. DeMali, A. L. Jue, and K. Burridge, “IpaA targets β1 integrins and rho to promote actin cytoskeleton rearrangements necessary for Shigella entry,” *J. Biol. Chem.*, vol. 281, no. 51, pp. 39534–39541, 2006.

[121] A. Skoudy, J. Mounier, A. Aruffo, H. Ohayon, P. Gounon, P. Sansonetti, and G. Tran Van Nhieu, “CD44 binds to the Shigella IpaB protein and participates in bacterial invasion of epithelial cells,” *Cell. Microbiol.*, vol. 2, no. 1, pp. 19–33, 2000.

[122] M. Watarai, S. Funato, and C. Sasakawa, “Interaction of Ipa proteins of Shigella flexneri with alpha5beta1 integrin promotes entry of the bacteria into mammalian cells.,” *J. Exp. Med.*, vol. 183, no. 3, pp. 991–999, 1996.

[123] H. Hilbi, J. E. Moss, D. Hersh, Y. Chen, J. Arondel, S. Banerjee, R. A. Flavell, J. Yuan, P. J. Sansonetti, and A. Zychlinsky, “Shigella-induced apoptosis is dependent on caspase-1 which binds to IpaB,” *J. Biol. Chem.*, vol. 273, no. 49, pp. 32895–32900, 1998.

[124] N. High, J. Mounier, M. C. Prevost, and P. J. Sansonetti, “IpaB of Shigella flexneri causes entry into epithelial cells and escape from the phagocytic vacuole.,” *EMBO J.*, vol. 11, no. 5, p. 1991, 1992.

[125] G. Tran Van Nhieu, E. Caron, A. Hall, and P. J. Sansonetti, “IpaC induces actin polymerization and filopodia formation during Shigella entry into epithelial cells,” *EMBO J.*, vol. 18, no. 12, pp. 3249–3262, 1999.

[126] N. M. Alto, F. Shao, C. S. Lazar, R. L. Brost, G. Chua, S. Mattoo, S. A. McMahon, P. Ghosh, T. R. Hughes, C. Boone, and others, “Identification of a bacterial type III effector family with G protein mimicry functions,” *Cell*, vol. 124, no. 1, pp. 133–145, 2006.

[127] K. Niebuhr, S. Giuriato, T. Pedron, D. J. Philpott, F. Gaits, J. Sable, M. P. Sheetz, C. Parsot, P. J. Sansonetti, and B. Payrastre, “Conversion of PtdIns (4, 5) P2 into PtdIns (5) P by the S. flexneri effector IpgD reorganizes host cell morphology,” *EMBO J.*, vol. 21, no. 19, pp. 5069–5078, 2002.

[128] S. Yoshida, E. Katayama, A. Kuwae, H. Mimuro, T. Suzuki, and C. Sasakawa, “Shigella deliver an effector protein to trigger host microtubule destabilization, which promotes Rac1 activity and efficient bacterial internalization,” *EMBO J.*, vol. 21, no. 12, pp. 2923–2935, 2002.

[129] S. Yoshida, Y. Handa, T. Suzuki, M. Ogawa, M. Suzuki, A. Tamai, A. Abe, E. Katayama, and C. Sasakawa, “Microtubule-severing activity of Shigella is pivotal for intercellular spreading,” *Science (80-.).*, vol. 314, no. 5801, pp. 985–989, 2006.

[130] K. L. Germane, R. Ohi, M. B. Goldberg, and B. W. Spiller, “Structural and Functional Studies Indicate That Shigella VirA Is Not a Protease and Does Not Directly Destabilize Microtubules‡,” *Biochemistry*, vol. 47, no. 39, pp. 10241–10243, 2008.

[131] M. Ogawa, T. Suzuki, I. Tatsuno, H. Abe, and C. Sasakawa, “IcsB, secreted via the type III secretion system, is chaperoned by IpgA and required at the post-invasion stage of Shigella pathogenicity,” *Mol. Microbiol.*, vol. 48, no. 4, pp. 913–931, 2003.

[132] T. Suzuki, M.-C. Lett, and C. Sasakawa, “Extracellular transport of VirG protein in Shigella,” *J. Biol. Chem.*, vol. 270, no. 52, pp. 30874–30880, 1995.

[133] K. Kuhnel and D. Diezmann, “Crystal structure of the autochaperone region from the Shigella flexneri autotransporter IcsA,” *J. Bacteriol.*, p. JB–00790, 2011.

[134] I. R. Henderson, F. Navarro-Garcia, M. Desvaux, R. C. Fernandez, and D. Ala’Aldeen, “Type V protein secretion pathway: the autotransporter story,” *Microbiol. Mol. Biol. Rev.*, vol. 68, no. 4, pp. 692–744, 2004.

[135] M. Y. Teh and R. Morona, “Identification of Shigella flexneri IcsA residues affecting interaction with N-WASP, and evidence for IcsA-IcsA co-operative interaction,” *PLoS One*, vol. 8, no. 2, p. e55152, 2013.

[136] M. Grabowicz, “Biogenesis of Shigella flexneri IcsA protein.,” 2010.

[137] S. Jain, P. van Ulsen, I. Benz, M. A. Schmidt, R. Fernandez, J. Tommassen, and M. B. Goldberg, “Polar localization of the autotransporter family of large bacterial virulence proteins,” *J. Bacteriol.*, vol. 188, no. 13, pp. 4841–4850, 2006.

[138] A. Janakiraman, K. R. Fixen, A. N. Gray, H. Niki, and M. B. Goldberg, “A genome-scale proteomic screen identifies a role for DnaK in chaperoning of polar autotransporters in Shigella,” *J. Bacteriol.*, vol. 191, no. 20, pp. 6300–6311, 2009.

[139] L. D. Brandon, N. Goehring, A. Janakiraman, A. W. Yan, T. Wu, J. Beckwith, and M. B. Goldberg, “IcsA, a polarly localized autotransporter with an atypical signal peptide, uses the Sec apparatus for secretion, although the Sec apparatus is circumferentially distributed,” *Mol. Microbiol.*, vol. 50, no. 1, pp. 45–60, 2003.

[140] C. Egile, H. D’Hauteville, C. Parsot, and P. J. Sansonetti, “SopA, the outer membrane protease responsible for polar localization of IcsA in Shigella flexneri,” *Mol. Microbiol.*, vol. 23, no. 5, pp. 1063–1073, 1997.

[141] K. Rajakumar, B. H. Jost, C. Sasakawa, N. Okada, M. Yoshikawa, and B. Adler, “Nucleotide sequence of the rhamnose biosynthetic operon of Shigella flexneri 2a and role of lipopolysaccharide in virulence.,” *J. Bacteriol.*, vol. 176, no. 8, pp. 2362–2373, 1994.

[142] R. C. Sandlin, K. A. Lampel, S. P. Keasler, M. B. Goldberg, A. L. Stolzer, and A. T. Maurelli, "Avirulence of rough mutants of Shigella flexneri: requirement of O antigen for correct unipolar localization of IcsA in the bacterial outer membrane.," *Infect. Immun.*, vol. 63, no. 1, pp. 229–237, 1995.

[143] D. C. Oliver, G. Huang, E. Nodel, S. Pleasance, and R. C. Fernandez, "A conserved region within the Bordetella pertussis autotransporter BrkA is necessary for folding of its passenger domain," *Mol. Microbiol.*, vol. 47, no. 5, pp. 1367–1383, 2003.

[144] M. B. Goldberg, O. Barzu, C. Parsot, and P. J. Sansonetti, "Unipolar localization and ATPase activity of IcsA, a Shigella flexneri protein involved in intracellular movement.," *J. Bacteriol.*, vol. 175, no. 8, pp. 2189–2196, 1993.

[145] T. Suzuki, H. Miki, T. Takenawa, and C. Sasakawa, "Neural Wiskott--Aldrich syndrome protein is implicated in the actin-based motility of Shigella flexneri," *EMBO J.*, vol. 17, no. 10, pp. 2767–2776, 1998.

[146] K. L. May and R. Morona, "Mutagenesis of the Shigella flexneri autotransporter IcsA reveals novel functional regions involved in IcsA biogenesis and recruitment of host neural Wiscott-Aldrich syndrome protein," *J. Bacteriol.*, vol. 190, no. 13, p. 4666, 2008.

[147] T. Suzuki, S. Saga, and C. Sasakawa, "Functional analysis of Shigella VirG domains essential for interaction with vinculin and actin-based motility," *J. Biol. Chem.*, vol. 271, no. 36, pp. 21878–21885, 1996.

[148] T. Chakraborty, F. Ebel, E. Domann, K. Niebuhr, B. Gerstel, S. Pistor, C. J. Temm-Grove, B. M. Jockusch, M. Reinhard, U. Walter, and others, "A focal adhesion factor directly linking intracellularly motile Listeria monocytogenes and Listeria ivanovii to the actin-based cytoskeleton of mammalian cells.," *EMBO J.*, vol. 14, no. 7, p. 1314, 1995.

[149] T. Suzuki and C. Sasakawa, "Molecular basis of the intracellular spreading of Shigella," *Infect. Immun.*, vol. 69, no. 10, pp. 5959–5966, 2001.

[150] M. B. Goldberg, "Shigella actin-based motility in the absence of vinculin," *Cell Motil. Cytoskeleton*, vol. 37, no. 1, pp. 44–53, 1997.

[151] Y. Wang, G.-H. Gong, W. Zhou, B. Zhang, S.-Y. Bao, C.-X. Wei, J.-J. Yue, and Y.-F. Zhang, "Analysis on the Interaction Domain of VirG and Apyrase by Pull-Down Assay," *Molecules*, vol. 19, no. 11, pp. 18090–18101, 2014.

[152] M. B. Goldberg and J. A. Theriot, "Shigella flexneri surface protein IcsA is sufficient to direct actin-based motility," *Proc. Natl. Acad. Sci.*, vol. 92, no. 14, pp. 6572–6576, 1995.

[153] J. Pei and N. V Grishin, "The Rho GTPase inactivation domain in Vibrio cholerae MARTX toxin has a circularly permuted papain-like thiol protease fold," *Proteins Struct. Funct. Bioinforma.*, vol. 77, no. 2, pp. 413–419, 2009.

[154] C. A. Kayath, S. Hussey, and others, "Escape of intracellular Shigella from autophagy requires binding to cholesterol through the type III effector, IcsB," *Microbes Infect.*, vol. 12, no. 12–13, pp. 956–966, 2010.

[155] U. Consortium and others, "UniProt: a hub for protein information," *Nucleic Acids Res.*, p. gku989, 2014.

[156] A. Lupas, M. Van Dyke, J. Stock, and others, “Predicting coiled coils from protein sequences,” *Science (80-.).*, vol. 252, no. 5009, pp. 1162–1164, 1991.

[157] F. S. Cordes, K. Komoriya, E. Larquet, S. Yang, E. H. Egelman, A. Blocker, and S. M. Lea, “Helical structure of the needle of the type III secretion system of Shigella flexneri,” *J. Biol. Chem.*, vol. 278, no. 19, pp. 17103–17107, 2003.

[158] M. F. Feldman and G. R. Cornelis, “The multitalented type III chaperones: all you can do with 15 kDa,” *FEMS Microbiol. Lett.*, vol. 219, no. 2, pp. 151–158, 2003.

[159] L. A. Baxt and M. B. Goldberg, “Host and Bacterial Proteins That Repress Recruitment of LC3 to Shigella Early during Infection,” *PLoS One*, vol. 9, no. 4, p. e94653, 2014.

[160] Y. Leung, S. Ally, and M. B. Goldberg, “Bacterial actin assembly requires toca-1 to relieve N-wasp autoinhibition,” *Cell Host Microbe*, vol. 3, no. 1, pp. 39–47, 2008.

[161] M. Matsushita, N. N. Suzuki, K. Obara, Y. Fujioka, Y. Ohsumi, and F. Inagaki, “Structure of Atg5{\textperiodcentered} Atg16, a Complex Essential for Autophagy,” *J. Biol. Chem.*, vol. 282, no. 9, p. 6763, 2007.

[162] E. Keil, R. Höcker, M. Schuster, F. Essmann, N. Ueffing, B. Hoffman, D. A. Liebermann, K. Pfeffer, K. Schulze-Osthoff, and I. Schmitz, “Phosphorylation of Atg5 by the Gadd45β--MEKK4-p38 pathway inhibits autophagy,” *Cell Death Differ.*, 2012.

[163] S. Yousefi, R. Perozzo, I. Schmid, A. Ziemiecki, T. Schaffner, L. Scapozza, T. Brunner, and H.-U. Simon, “Calpain-mediated cleavage of Atg5 switches autophagy to apoptosis,” *Nat. Cell Biol.*, vol. 8, no. 10, pp. 1124–1132, 2006.

[164] A. Williams, S. Sarkar, P. Cuddon, E. K. Ttofi, S. Saiki, F. H. Siddiqi, L. Jahreiss, A. Fleming, D. Pask, P. Goldsmith, and others, “Novel targets for Huntington’s disease in an mTOR-independent autophagy pathway,” *Nat. Chem. Biol.*, vol. 4, no. 5, pp. 295–305, 2008.

[165] J. H. Kim, S. B. Hong, J. K. Lee, S. Han, K.-H. Roh, K.-E. Lee, Y. K. Kim, E.-J. Choi, and H. K. Song, “Insights into autophagosome maturation revealed by the structures of ATG5 with its interacting partners,” *Autophagy*, no. just-accepted, p. 0, 2014.

[166] D. Chen, W. Fan, Y. Lu, X. Ding, S. Chen, and Q. Zhong, “A mammalian autophagosome maturation mechanism mediated by TECPR1 and the Atg12-Atg5 conjugate,” *Mol. Cell*, vol. 45, no. 5, pp. 629–641, 2012.

[167] A. Ashkenazi, V. M. Dixit, and others, “Death receptors: signaling and modulation,” *Science (80-.).*, vol. 281, no. 5381, pp. 1305–1308, 1998.

[168] T. S. Walter, C. Meier, R. Assenberg, K.-F. Au, J. Ren, A. Verma, J. E. Nettleship, R. J. Owens, D. I. Stuart, and J. M. Grimes, “Lysine methylation as a routine rescue strategy for protein crystallization,” *Structure*, vol. 14, no. 11, pp. 1617–1622, 2006.

[169] M. V Petoukhov, D. Franke, A. V Shkumatov, G. Tria, A. G. Kikhney, M. Gajda, C. Gorba, H. D. T. Mertens, P. V Konarev, and D. I. Svergun, “New developments in the ATSAS program package for small-angle scattering data analysis,” *J. Appl. Crystallogr.*, vol. 45, no. 2, pp. 342–350, 2012.

[170] P. V Konarev, V. V Volkov, A. V Sokolova, M. H. J. Koch, and D. I. Svergun, “PRIMUS: a Windows PC-based system for small-angle scattering data analysis,” *J. Appl. Crystallogr.*, vol. 36, no. 5, pp. 1277–1282, 2003.

[171] D. Svergun, C. Barberato, and M. H. J. Koch, “CRYSOL--a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates,” *J. Appl. Crystallogr.*, vol. 28, no. 6, pp. 768–773, 1995.

[172] D. I. Svergun, “Determination of the regularization parameter in indirect-transform methods using perceptual criteria,” *J. Appl. Crystallogr.*, vol. 25, no. 4, pp. 495–503, 1992.

[173] D. Franke and D. I. Svergun, “DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering,” *J. Appl. Crystallogr.*, vol. 42, no. 2, pp. 342–346, 2009.

[174] V. V Volkov and D. I. Svergun, “Uniqueness of ab initio shape determination in small-angle scattering,” *J. Appl. Crystallogr.*, vol. 36, no. 3, pp. 860–864, 2003.

[175] M. V Petoukhov and D. I. Svergun, “Global rigid body modeling of macromolecular complexes against small-angle scattering data,” *Biophys. J.*, vol. 89, no. 2, pp. 1237–1250, 2005.

[176] H. Fischer, M. de Oliveira Neto, H. B. Napolitano, I. Polikarpov, and A. F. Craievich, “Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale,” *J. Appl. Crystallogr.*, vol. 43, no. 1, pp. 101–109, 2009.

[177] B. Rupp, *Biomolecular crystallography: principles, practice, and application to structural biology*. Garland Science, 2009.

[178] W. Kabsch, “Xds,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 125–132, 2010.

[179] P. Evans, “Scaling and assessment of data quality,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 62, no. 1, pp. 72–82, 2006.

[180] P. D. Adams, P. V Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, and others, “PHENIX: a comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 213–221, 2010.

[181] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.

[182] T. C. Terwilliger, R. W. Grosse-Kunstleve, P. V Afonine, N. W. Moriarty, P. H. Zwart, L.-W. Hung, R. J. Read, and P. D. Adams, “Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 64, no. 1, pp. 61–69, 2008.

[183] P. Emsley and K. Cowtan, “Coot: model-building tools for molecular graphics,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12, pp. 2126–2132, 2004.

[184] P. V Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, and P. D. Adams, “Towards

automated crystallographic structure refinement with phenix. refine," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 68, no. 4, pp. 352–367, 2012.

[185] H. Yumerefendi, F. Tarendeau, P. J. Mas, and D. J. Hart, "ESPRIT: an automated, library-based method for mapping and soluble expression of protein domains from challenging targets," *J. Struct. Biol.*, vol. 172, no. 1, pp. 66–74, 2010.

[186] N. R. Movva, K. Nakamura, and M. Inouye, "Amino acid sequence of the signal peptide of ompA protein, a major outer membrane protein of Escherichia coli.," *J. Biol. Chem.*, vol. 255, no. 1, pp. 27–29, 1980.

[187] D. J. Leibly, T. N. Nguyen, L. T. Kao, S. N. Hewitt, L. K. Barrett, and W. C. Van Voorhis, "Stabilizing Additives Added during Cell Lysis Aid in the Solubilization of Recombinant Proteins," *PLoS One*, vol. 7, no. 12, p. e52482, 2012.

[188] J. M. Song, Y. J. An, M. H. Kang, Y. H. Lee, and S. S. Cha, "Cultivation at 6--10° C is an effective strategy to overcome the insolubility of recombinant proteins in Escherichia coli," *Protein Expr. Purif.*, vol. 82, no. 2, pp. 297–301, 2012.

[189] K. L. Longenecker, S. M. Garrard, P. J. Sheffield, and Z. S. Derewenda, "Protein crystallization by rational mutagenesis of surface residues: Lys to Ala mutations promote crystallization of RhoGDI," *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 57, no. 5, pp. 679–688, 2001.

[190] L. Goldschmidt, D. R. Cooper, Z. S. Derewenda, and D. Eisenberg, "Toward rational protein crystallization: A Web server for the design of crystallizable protein variants," *Protein Sci.*, vol. 16, no. 8, pp. 1569–1576, 2007.

[191] A. Dong, X. Xu, A. M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, and others, "In situ proteolysis for protein crystallization and structure determination," *Nat. Methods*, vol. 4, no. 12, pp. 1019–1021, 2007.

[192] N. J. Greenfield, "Using circular dichroism spectra to estimate protein secondary structure," *Nat. Protoc.*, vol. 1, no. 6, pp. 2876–2890, 2006.

[193] J. Reed and T. A. Reed, "A set of constructed type spectra for the practical estimation of peptide secondary structure from circular dichroism," *Anal. Biochem.*, vol. 254, no. 1, pp. 36–40, 1997.

[194] T. Suzuki, H. Mimuro, S. Suetsugu, H. Miki, T. Takenawa, and C. Sasakawa, "Neural Wiskott--Aldrich syndrome protein (N-WASP) is the specific ligand for Shigella VirG among the WASP family and determines the host cell type allowing actin-based spreading," *Cell. Microbiol.*, vol. 4, no. 4, pp. 223–233, 2002.

[195] D. W. Leung, D. M. Morgan, and M. K. Rosen, "Biochemical Properties and Inhibitors of (N-) WASP," *Methods Enzymol.*, vol. 406, pp. 281–296, 2006.

[196] R. D. Hayward, R. J. Cain, E. J. McGhie, N. Phillips, M. J. Garner, and V. Koronakis, "Cholesterol binding by the bacterial type III translocon is essential for virulence effector delivery into mammalian cells," *Mol. Microbiol.*, vol. 56, no. 3, pp. 590–603, 2005.

[197] L. A. Kelley and M. J. E. Sternberg, "Protein structure prediction on the Web: a case study using the Phyre server," *Nat. Protoc.*, vol. 4, no. 3, pp. 363–371, 2009.

[198] H.-C. Cheng, B. M. Skehan, K. G. Campellone, J. M. Leong, and M. K. Rosen, “Structural mechanism of WASP activation by the enterohaemorrhagic E. coli effector EspFu,” *Nature*, vol. 454, no. 7207, pp. 1009–1013, 2008.

[199] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon, “Electrostatics of nanosystems: application to microtubules and the ribosome,” *Proc. Natl. Acad. Sci.*, vol. 98, no. 18, pp. 10037–10041, 2001.

[200] T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe, and N. A. Baker, “PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations,” *Nucleic Acids Res.*, vol. 35, no. suppl 2, pp. W522–W525, 2007.

[201] K. Hsin, Y. Sheng, M. M. Harding, Taylor P, and M. D. and Walkinshaw, “MESPEUS: a database of the geometry of metal sites in proteins,” *J. Appl. Crystallogr.*, vol. 41, no. 5, pp. 963–968, 2008.

[202] E. Krissinel and K. Henrick, “Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute.” 2007.

[203] L. Holm and P. Rosenström, “Dali server: conservation mapping in 3D,” *Nucleic Acids Res.*, vol. 38, no. suppl 2, pp. W545–W549, 2010.

[204] E. Krissinel and K. Henrick, “Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12, pp. 2256–2268, 2004.

[205] C. Anna and J. Behrens, “Crystal Structure & Characterization of the SCOC Coiled Coil Domain submitted by,” 2013.

[206] Y. Fu and J. E. Galán, “Identification of a specific chaperone for SptP, a substrate of the centisome 63 type III secretion system ofSalmonella typhimurium,” *J. Bacteriol.*, vol. 180, no. 13, pp. 3393–3399, 1998.

[207] P. A. Bronstein, E. A. Miao, and S. I. Miller, “InvB is a type III secretion chaperone specific for SspA,” *J. Bacteriol.*, vol. 182, no. 23, pp. 6638–6644, 2000.

[208] C. E. Stebbins and J. E. Galán, “Priming virulence factors for delivery into the host,” *Nat. Rev. Mol. Cell Biol.*, vol. 4, no. 9, pp. 738–744, 2003.

[209] M. Pagni, V. Ioannidis, L. Cerutti, M. Zahn-Zabal, C. V. Jongeneel, J. Hau, O. Martin, D. Kuznetsov, and L. Falquet, “MyHits: improvements to an interactive resource for analyzing protein sequences,” *Nucleic Acids Res.*, vol. 35, no. suppl 2, pp. W433–W437, 2007.

[210] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, and others, “Pfam: the protein families database,” *Nucleic Acids Res.*, p. gkt1223, 2013.

[211] D. W. A. Buchan, F. Minneci, T. C. O. Nugent, K. Bryson, and D. T. Jones, “Scalable web services for the PSIPRED Protein Analysis Workbench,” *Nucleic Acids Res.*, vol. 41, no. W1, pp. W349–W357, 2013.

[212] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, and D. T. Jones, “Prediction and functional analysis of native disorder in proteins from the three kingdoms of life,” *J. Mol. Biol.*, vol. 337, no. 3, pp. 635–645, 2004.

[213] O. Aitio, M. Hellman, B. Skehan, T. Kesti, J. M. Leong, K. Saksela, and P. Permi, "Enterohaemorrhagic Escherichia Coli Exploits a Tryptophan Switch to Hijack Host F-Actin Assembly," *Structure*, 2012.

[214] Z. Metlagel, C. Otomo, G. Takaesu, and T. Otomo, "Structural basis of ATG3 recognition by the autophagic ubiquitin-like protein ATG12," *Proc. Natl. Acad. Sci.*, vol. 110, no. 47, pp. 18844–18849, 2013.

[215] C. Otomo, Z. Metlagel, G. Takaesu, and T. Otomo, "Structure of the human ATG12~ ATG5 conjugate required for LC3 lipidation in autophagy," *Nat. Struct. Mol. Biol.*, 2012.

[216] S. Mostowy, M. Bonazzi, M. A. Hamon, T. N. Tham, A. Mallet, M. Lelek, E. Gouin, C. Demangel, R. Brosch, C. Zimmer, and others, "Entrapment of intracytosolic bacteria by septin cage-like structures," *Cell Host Microbe*, vol. 8, no. 5, pp. 433–444, 2010.

[217] L. Biskri and A. Allaoui, "Shigella Stands up to the Challenge of Adhesion," *Cell Host Microbe*, vol. 15, no. 4, pp. 393–395, 2014.

Appendix

A1. Oligonucleotides

Table 22: Sequencing primers used in this work.

Description	Identifier	Sequence
T7 forward	SA016	taatacgactcactataggg
T7 reverse	SA017	gctagttattgctcagcgg
Duet mcs1 forward	PV001	ggatctcgacgctctccct
Duet mcs1 reverse	PV002	gattatgcggccgtgtacaa
Duet mcs2 forward	PV003	ttgtacacggccgcataatc
ESPRIT reverse	PV023	cctcaagacccgtttagaggcc
pGEX forward	MOSB23	atagcatggcctttgcagg
pGEX reverse	MOSB24	gagctgcatgtgtcagagg
VirG G159 forward	LS001	gtgctgaccataacggtgatgg
LS047	pESPRIT_seq_for	ggaaaaacgccagcaacgcg
LS048	pESPRIT_seq_rev	gccatcctatggaactgcctcgg

Table 23: VirG cloning primers used in this work.

Description	Identifier	Sequence
VirG M1 forward	SA009	AAGAATGCGGCCGCatgaatcaaattcacaaatttttttg
VirG M1 forward	LS038	ATGCATGCGGCGCGCCgtatgaatcaaattcacaaatttttttg
VirG T53 forward	SA010	AAGAATGCGGCCGCactcctctttcgggtactcaag
VirG G77 forward	LS092	AAGAATGCGGCCGCggactttctcccttgggagctg
VirG G117 forward	LS133	AAGAATGCGGCCGCggtcatggtggtgctggtg
VirG G131 forward	LS134	AAGAATGCGGCCGCggcggtaatggtggtgactc
VirG N281 forward	LS135	AAGAATGCGGCCGCaacaatggagctattttaggcgg
VirG Q320 forward	SA011	AAGAATGCGGCCGCcaaaatgtagcaggtaatgctatcc
VirG T303 forward	LS094	AAGAATGCGGCCGCaccattgctaatagcggatatataatttc
VirG E340 forward	LS093	AAGAATGCGGCCGCgaaggttctgtcattactggtgatg
VirG G341 forward	LS083	AAGAATGCGGCCGCggttctgtcattactggtgatgtac
VirG Q349 forward	LS150	AAGAATGCGGCCGCcaggttaacaattcatccattctg
VirG V350 forward	LS082	AAGAATGCGGCCGCgttaacaattcatccattctgaaaattatc
VirG S353 forward	LS089	AAGAATGCGGCCGCtcatccattctgaaaattatcaacaatg
VirG S353 forward	LS152	AAGAATGCGGCCGCtcatccattctgaaaattatcaacaatg
VirG N361 forward	LS085	AAGAATGCGGCCGCaatgattacactgggaccacacc
VirG Y363 forward	LS084	AAGAATGCGGCCGCtacactgggaccacaccaactattg
VirG T364 forward	LS087	AAGAATGCGGCCGCactgggaccacaccaactattg
VirG G377 forward	LS080	AAGAATGCGGCCGCggtgattgtacaactgtttcactatc
VirG G377 forward	LS153	AAGAATGCGGCCGCggtgattgtacaactgtttcactatc
VirG S383 forward	LS151	AAGAATGCGGCCGCtcactatcaggtaacaaattcactg
VirG G398 forward	LS154	AAGAATGCGGCCGCggtgagaacagttctttaaatttag
VirG S409 forward	LS086	AAGAATGCGGCCGCagtagtctggaagcttctggaaatatg

VirG F419 forward	LS096	AAGAATGCGGCCGCtttggcaacaatgtaaaagtggag
VirG V425 forward	LS081	AAGAATGCGGCCGCgtggaggctattataaataactgg
VirG N430 forward	LS088	AAGAATGCGGCCGCaataactgggcgcagaaggac
VirG A433 forward	LS079	AAGAATGCGGCCGCgcgcagaaggactataaactgc
VirG T469 forward	LS090	AAGAATGCGGCCGCacaaaaagctatatcagtgaccag
VirG S526 forward	LS095	AAGAATGCGGCCGCtccctaacaaaatcaggggagg
VirG N577 forward	LS091	AAGAATGCGGCCGCaatttttcaggcatgaaccaaactg
VirG G420 reverse	LS136	GGGGTACCttagccaaatgacatatttccagaagc
VirG A433 reverse	SA013	GGGGTACCtcacgcccagttatttataatagcctc
VirG S484 reverse	LS137	GGGGTACCttagctcaaaccgtagatcaatttattc
VirG G592 reverse	SA048	GGGGTACCttacccactatttaataaagtgttaacag
VirG G592 reverse	LS040	ATGCATGCATGCATtcccactatttaataaagtgttaacag
VirG D740 reverse	LS164	GGGGTACCttaatccatctgactagttagataccac
VirG R758 reverse	SA014	GGGGTACCtcagcgactactcatttgagtagactc
VirG R758 reverse	LS039	ATGCATGCATGCATtgcgactactcatttgagtagactc
VirG R758 reverse	LS097	GGGGTACCttagcgactactcatttgagtagactc
VirG R779 reverse	SA015	GGGGTACCtcagcgccatgtgtgaatactacc
VirG F1102 reverse	SA012	GGGGTACCtcagaaggtatatttcacaccc

Table 24: IcsB and IpgA cloning primers used in this work.

Description	Identifier	Sequence
IcsB M1 forward	SA001	AAGAATGCGGCCGCatgatcctcaaaattagcaatttcattg
IcsB M1 forward	LS041	ATGCATGCGGATCCatgatcctcaaaattagcaatttcattg
IcsB S6 forward	LS049	AAGAATGCGGCCGCagcaatttcattgacgcaagcaatac
IcsB F8 forward	LS050	AAGAATGCGGCCGCttcattgacgcaagcaatacaaaag
IcsB G16 forward	LS002	AAGAATGCGGCCGCgggcctatacgcgttgaagatac
IcsB G26 forward	LS003	AAGAATGCGGCCGCggacctatattgatagcacagaaattc
IcsB G121 forward	LS163	AAGAATGCGGCCGCggtacagccaaaaatcttgaaagac
IcsB G169 forward	LS004	AAGAATGCGGCCGCgggaatgttacaaaagatacag
IcsB I177 forward	LS051	AAGAATGCGGCCGCataaactacctttttgaaaaacgctc
IcsB S185 forward	LS005	AAGAATGCGGCCGCtcaggttacagtgtcgatacc
IcsB S188 forward	LS052	AAGAATGCGGCCGCagtgtcgatacctataagcaagac
IcsB V214 forward	LS053	AAGAATGCGGCCGCgttagaaacttattaaattcaaaacaag
IcsB P233 forward	LS054	AAGAATGCGGCCGCcccagagctaatcaaaaaaaagac
IcsB G265 forward	LS006	AAGAATGCGGCCGCggtaaaatcagccataacctttttgg
IcsB N125 reverse	LS012	GGGGTACCttaatttttggctgtaccgatgccatg
IcsB L130 reverse	LS103	GGGGTACCttataaatgtctttcaagatttttggctg
IcsB G136 reverse	LS105	GGGGTACCttatccccagcaacgaaaaactaaatg
IcsB P167 reverse	LS055	GGGGTACCttagggataccagctgagataggtatg
IcsB E176 reverse	LS056	GGGGTACCttattctgtatcttttgtaacattcccg
IcsB K183 reverse	LS057	GGGGTACCttatttttcaaaaaggtagtttatttctgtatc
IcsB K229 reverse	LS013	GGGGTACCttatttttgttgttattttgatcttgttttg
IcsB P391 reverse	LS058	GGGGTACCttatggtgtgtggataaggttaattttattg

IcsB S428 reverse	LS059	GGGGTACCttatgattttactaagtatgagtggcaag
IcsB N438 reverse	LS060	GGGGTACCttagtttaatacacgagtcatattgctg
IcsB S464 reverse	LS061	GGGGTACCttaagaattgtaataatcaataatttttattg
IcsB N493 reverse	LS062	GGGGTACCttaattagaatgagagttattcaataaaag
IcsB I494 reverse	LS007	GGGGTACCttatatattagaatgagagttattcaataaaag
IcsB I494 reverse	SA003	GGGGTACCctatatattagaatgagagttattcaataaaag
IcsB I494 reverse	LS042	ATGCATGCGCGGCCGCttatatattagaatgagagttattcaataaaag
IpgA M1 forward	SA007	AAGAATGCGGCCGCatgtgtcgcaaactatatgataaac
IpgA M1 forward	LS022	GGGGTTTCATatgtgtcgcaaactatatgataaac
IpgA N129 reverse	SA008	GGGGTACCttagttcacttctgaagtgatgtttg
IpgA N129 reverse	LS023	GGAATTCttagttcacttctgaagtgatgtttg

Table 25: ATG5 cloning primers used in this work.

Description	Identifier	Sequence
ATG5 M1 forward	SA033	AAGAATGCGGCCGCatgacagatgacaaagatgtgcttc
ATG5 D275 reverse	SA034	GGGGTACCtcaatctgttggctgtgggatg

Table 26: Other cloning primers used in this work.

Description	Identifier	Sequence
ATG16L1 M1 forward	SA018	AAGAATGCGGCCGCatgtcgtcgggcctccgcgc
ATG16L1 L43 C reverse	LS045	GGGGTACCttagcacagcaatttgttatactgcaggatg
ATG16L1 L43 reverse	LS046	GGGGTACCttacagcaatttgttatactgcaggatg
WASP S242 forward	LS071	AAGAATGCGGCCGCagtggattcaagcatgtcagcc
WASP Q461 forward	LS073	AAGAATGCGGCCGCcagagctcagagggactggtg
WASP Q310 reverse	LS072	GGGGTACCttactggcgcctcatctcctgcc
WASP G492 reverse	LS074	GGGGTACCttagccagcctggtcctccccttc
N-WASP S207 forward	LS075	AAGAATGCGGCCGCagcaatttccagcacattggacatg
N-WASP V399 forward	LS106	AAGAATGCGGCCGCgttccaactactgcaggaaac
N-WASP A462 forward	LS077	AAGAATGCGGCCGCgcacccacttcaggaattgtgg
N-WASP Q275 reverse	LS076	GGGGTACCttattgcctccgcagttcatttttaac
N-WASP T460 reverse	LS108	GGGGTACCttatgttggtggtgtagactcttgg
N-WASP D493 reverse	LS078	GGGGTACCttaatcatcttcatcttcatcttcatctg
N-WASP D505 reverse	LS107	GGGGTACCttagtcttcccactcatcatcatcc
Toca-1 M1 forward	LS143	AAGAATGCGGCCGCatgagttggggcacggaactg
Toca-1 S245 forward	LS144	AAGAATGCGGCCGCtcgaagtgtctggagggtatg
Toca-1 G383 forward	LS110	AAGAATGCGGCCGCggtgtctctctgattatgggcc
Toca-1 G389 forward	LS109	AAGAATGCGGCCGCggccctgcgctggaagatttc
Toca-1 T293 reverse	LS145	GGGGTACCttaggtgcggtagatatgctggctg
Toca-1 M431 reverse	LS146	GGGGTACCttacattttgttaagagcatctttctgg
Toca-1 G480 reverse	LS111	GGGGTACCttaacccgttttgccttccacttc
Toca-1 S488 reverse	LS112	GGGGTACCttaggaatgacgccgatctccgcg
Toca-1 D535 reverse	LS147	GGGGTACCttagtcatcgtcctcaaactcatcatc

Cip4 S392 forward	LS113	AAGAATGCGGCCGCagccacttgcccccagagc
Cip4 G480 reverse	LS114	GGGGTACCttatccccggttgctaaggactcg
Cdc42 M1 forward	LS115	AAGAATGCGGCCGCatgcagacaattaagtgtgttgttg
Cdc42 E178 reverse	LS142	GGGGTACCttactccagggcagccaatattgc
Diaph1 D78 forward	LS117	AAGAATGCGGCCGCgatcccacagcacagtcattgc
Diaph1 E460 reverse	LS118	GGGGTACCttactcaatctggaggtgccggc
Diaph2 P88 forward	LS119	AAGAATGCGGCCGCcctgcagctcagccattatatg
Diaph2 D476 reverse	LS120	GGGGTACCttaatcgatgtctaatctttgcctg
Diaph3 M264 forward	LS121	AAGAATGCGGCCGCatgagtgaggagaggagcctttc
Diaph3 D488 reverse	LS122	GGGGTACCttaatctaaatctagtctttttcgatatg
Daam2 S39 forward	LS123	AAGAATGCGGCCGCtccagccccatcccgaacgc
Daam2 N427 reverse	LS124	GGGGTACCttaattgaagttctccaagggagcc
Fhod1 S44 forward	LS125	AAGAATGCGGCCGCagcctggacggggcgctgc
Fhod1 G468 reverse	LS126	GGGGTACCttacccggcaagtgtctctgccc
Fmnl1 A23 forward	LS127	AAGAATGCGGCCGCgcgcctcccaagcagccg
Fmnl1 A468 reverse	LS128	GGGGTACCttaagcgggaggcgctttctcag
Fmnl2 A16 forward	LS129	AAGAATGCGGCCGCgcgcacaacgtgcctttgaag
Fmnl2 G487 reverse	LS130	GGGGTACCttacccttttcttctgaattttaatggtc
Fmnl3 S16 forward	LS131	AAGAATGCGGCCGCtctgtcccgttgttgctgccg
Fmnl3 G477 reverse	LS132	GGGGTACCttagccccggacatttggctcc

Table 27: Mutagenesis primers used in this work.

Description	Identifier	Sequence
VirG NsiI out forward	LS043	tctcatgcgtctagaactaagctacggattaac
VirG NsiI out reverse	LS044	tctagacgcatgagaggggttggaagaag
VirG del forward	LS155	gtcagatacagatggcgacagtcatg
VirG del GGS forward	LS157	tcaggtggcagcgatacagatggcgacagtcatg
VirG del 2GGS forward	LS159	gtggcagcggtggcagcgatacagatggcgacagtcatg
VirG del 4GGS forward	LS161	ggtggcagcggtggcagcggtggcagcggtggcagcgatacagatggcgacagtcatg
VirG del reverse	LS156	ccatctgtatctgacatatttccagaagcttccag
VirG del GGS reverse	LS158	atcgctgccacctgacatatttccagaagcttccag
VirG del 2GGS reverse	LS160	ctgccaccgctgccacctgacatatttccagaagcttccag
VirG del 4GGS reverse	LS162	gctgccaccgctgccaccgctgccaccgctgccacctgacatatttccagaagcttccag
IcsB TAA forward	LS010	cattctaatatataaaaggccatagaaatg
IcsB TAA reverse	LS011	catttctatggcctttatatattagaatg
IcsB SER1 forward	LS024	aatgcggcagccgaccctatggttattgggggg
IcsB SER1 reverse	LS025	gtcggctgccgcattagctctgggaaaaaatatttttttg
IcsB SER2 forward	LS026	gatgcggcagccgaaatcatcgaaaaatcaataaaaattattg
IcsB SER2 reverse	LS027	ttcggctgccgcatctgttgcttcggtttttaattcg
IcsB SER3 forward	LS028	tgcgcggcagccgcagatgctttcagacaattggc
IcsB SER3 reverse	LS029	tgcggctgccgcgcatataaattttgacatgtttgtttc
IcsB SER4 forward	LS030	gcggcatctgcagccagtgcgtttgccgcctatgaag

IcsB SER4 reverse	LS031	ggctgcagatgccgcgctagctaatgtattgaggaaaagc
IcsB SER5 forward	LS032	tttgcggcacgctcaggttacagtgtcgatac
IcsB SER5 reverse	LS033	tgagcgtgccgcaaaaaggtagtttatttctgtatcttttg
IcsB SER6 forward	LS034	aacaacgcggcaatatttttttcccagagctaatcaaaaaa
IcsB SER6 reverse	LS035	tattgccgcgttgttattttgatcttgttttgaattta
IcsB GGS Q99 forward	LS104	gtggcagcggtggcagccaacactctttcacaggtagagac
IcsB GGS G104 forward	LS102	gtggcagcggtggcagcggtagagacatcaaattaatgttg
ATG5 T75E forward	LS036	gaaggcgaaccactgaaatggcattatccaattg
ATG5 T75E reverse	LS037	cagtggttcgccttcatattcaaaccatatctc
N-WASP Q275 GGS reverse	LS100	ctgccaccgctgccaccttgcctccgcagttcatttttaac
N-WASP GGS A462 forward	LS101	gtggcagcggtggcagcgcacccacttcaggaattgtgg
Toca-1 del forward	LS148	gaaaccgaaaggccctgcgctggaagatttc
Toca-1 del reverse	LS149	gggcctttcggtttcttcccgaaaagcc
Daam2 KpnI out forward	LS138	ctgggccgttaccgggatgaagtgaatctgaaaacagcc
Daam2 KpnI out reverse	LS139	tcccggtaacggcccagacttcggtctagctcgttc
Daam2 W414R forward	LS140	gatgagcgtggtgtggaccctgacctggctccc
Daam2 W414R reverse	LS141	cacaccacgctcatcctggaggacaatctgctggag

Table 28: Primers used in this work for annealing of oligonucleotides.

Description	Identifier	Sequence
IcsB Q99 forward	LS067	ggccgccaacactctttcacaggtagagacatcaaattaatgttgaataccgcagagcgttttcatgg
IcsB G136 forward	LS068	catcggtacagccaaaaatcttgaaagacatttagtttttcgttgctggggataaggtac
IcsB Q99 reverse	LS070	cggtattcaacattaatttgatgtctctacctgtgaaagagtgttggc
IcsB G136 reverse	LS069	cttatccccagcaacgaaaaactaaatgtctttcaagatttttggctgtaccgatgccatgaaaacgctctg
IcsB G104 forward	LS063	ggccgcggtagagacatcaaattaatgttgaataccgcagagcgttttcatgg
IcsB L130 forward	LS064	catcggtacagccaaaaatcttgaaagacatttataaggtac
IcsB G104 reverse	LS066	cggtattcaacattaatttgatgtctctaccgc
IcsB L130 reverse	LS065	cttataaatgtctttcaagatttttggctgtaccgatgccatgaaaacgctctg
OmpA 1 forward	LS014	catgaaaaagacagctatcgcgatt
OmpA 2 forward	LS015	gcagtggcactggctggtttcgctaccgtagcgcaggc
OmpA 3 reverse	LS016	catggcctgcgctacggtagcgaaac
OmpA 4 reverse	LS017	cagccagtgccactgcaatcgcgatagctgtcttttt
PelB 1 forward	LS018	catgaaatacctgctgccgaccgctgct
PelB 2 forward	LS019	gctggtctgctgctcctcgctgcccagccggcgatggc
PelB 3 reverse	LS020	catggccatcgccggctgggcagcgag
PelB 4 reverse	LS021	gagcagcagaccagcagcagcggtcggcagcaggtattt
mcs2 forward	LS098	tatggaattcg
mcs2 reverse	LS099	gatccgaattcca

A2. Protein constructs

Table 29: VirG protein constructs generated in this work.

Description	Identifier	Construct	Mutations
VirGK1	pLS004	Strep-TEV-VirG(53-433)	
VirGK1	pLS022	GST-Thr-TEV-VirG(53-433)	
VirGK1	pLS045	OmpA-Strep-TEV-VirG(53-433)	
VirGK2	pLS005	Strep-TEV-VirG(53-592)	
VirGK2	pLS023	GST-Thr-TEV-VirG(53-592)	
VirGK2	pLS046	OmpA-Strep-TEV-VirG(53-592)	
VirGK3	pLS006	Strep-TEV-VirG(53-758)	
VirGK3	pLS024	GST-Thr-TEV-VirG(53-758)	
VirGK3	pLS047	OmpA-Strep-TEV-VirG(53-758)	
VirGK4	pLS002	His-TEV-VirG(53-779)	
VirGK4	pLS007	Strep-TEV-VirG(53-779)	
VirGK4	pLS025	GST-Thr-TEV-VirG(53-779)	
VirGK4	pLS048	OmpA-Strep-TEV-VirG(53-779)	
VirGK5	pLS008	Strep-TEV-VirG(320-433)	
VirGK5	pLS026	GST-Thr-TEV-VirG(320-433)	
VirGK6	pLS009	Strep-TEV-VirG(320-592)	
VirGK6	pLS027	GST-Thr-TEV-VirG(320-592)	
VirGK7	pLS010	Strep-TEV-VirG(320-758)	
VirGK7	pLS028	GST-Thr-TEV-VirG(320-758)	
VirGK8	pLS011	Strep-TEV-VirG(320-779)	
VirGK8	pLS029	GST-Thr-TEV-VirG(320-779)	
VirGK9	pLS165	His-TEV-VirG(53-420)	
VirGK10	pLS166	His-TEV-VirG(77-420)	
VirGK11	pLS167	His-TEV-VirG(117-420)	
VirGK12	pLS168	His-TEV-VirG(131-420)	
VirGK13	pLS169	His-TEV-VirG(281-420)	
VirGK14	pLS170	His-TEV-VirG(53-484)	
VirGK15	pLS171	His-TEV-VirG(77-484)	
VirGK16	pLS172	His-TEV-VirG(117-484)	
VirGK17	pLS173	His-TEV-VirG(131-484)	
VirGK18	pLS174	His-TEV-VirG(281-484)	
VirGK19	pLS175	His-TEV-VirG(53-758)	
VirGK20	pLS176	His-TEV-VirG(77-758)	
VirGK21	pLS177	His-TEV-VirG(117-758)	
VirGK22	pLS178	His-TEV-VirG(131-758)	
VirGK23	pLS179	His-TEV-VirG(281-758)	
VirGK24	pLS184	His-TEV-VirG(340-758)	Δ419-486
VirGK25	pLS185	His-TEV-VirG(340-758)	Δ419-486 GGS
VirGK26	pLS186	His-TEV-VirG(340-758)	Δ419-486 2xGGS
VirGK27	pLS187	His-TEV-VirG(340-758)	Δ419-486 4xGGS
VirGK28	pLS188	His-TEV-VirG(349-758)	
VirGK29	pLS189	His-TEV-VirG(383-758)	

VirGK30	pLS190	His-TEV-VirG(353-758)	
VirGK31	pLS191	His-TEV-VirG(377-758)	
VirGK32	pLS192	His-TEV-VirG(398-758)	
VirGK33	pLS193	His-TEV-VirG(349-758)	Δ419-486 2xGGS
VirGK34	pLS194	His-TEV-VirG(383-758)	Δ419-486 2xGGS
VirGK35	pLS195	His-TEV-VirG(353-758)	Δ419-486 2xGGS
VirGK36	pLS196	His-TEV-VirG(377-758)	Δ419-486 2xGGS
VirGK37	pLS197	His-TEV-VirG(398-758)	Δ419-486 2xGGS
VirGK38	pLS198	His-TEV-VirG(340-740)	Δ419-486 2xGGS
VirGK39	pLS199	His-TEV-VirG(349-740)	Δ419-486 2xGGS
VirGK40	pLS200	His-TEV-VirG(383-740)	Δ419-486 2xGGS
VirGK41	pLS201	His-TEV-VirG(340-740)	
VirGK42	pLS202	His-TEV-VirG(349-740)	
VirGK43	pLS203	His-TEV-VirG(383-740)	
VirGK1/ ATG5 coexpression	pLS004/ pLS001	Strep-TEV-VirG(53-433)/ His-TEV-ATG5(1-275)	
VirGK1/ IcsB coexpression	pLS004/ pLS012	Strep-TEV-VirG(53-433)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK1/ ATG5 coexpression	pLS022/ pLS003	GST-Thr-TEV-VirG(53-433)/ Strep-TEV-ATG5(1-275)	
VirGK1/ IcsB coexpression	pLS022/ pLS013	GST-Thr-TEV-VirG(53-433)/ Strep-TEV-IcsB(1-494) IpgA(1-129)	
VirGK2/ ATG5 coexpression	pLS005/ pLS001	Strep-TEV-VirG(53-592)/ His-TEV-ATG5(1-275)	
VirGK2/ IcsB coexpression	pLS005/ pLS012	Strep-TEV-VirG(53-592)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK2/ IcsB coexpression	pLS023/ pLS013	GST-Thr-TEV-VirG(53-592)/ Strep-TEV-IcsB(1-494) IpgA(1-129)	
VirGK3/ ATG5 coexpression	pLS006/ pLS001	Strep-TEV-VirG(53-758)/ His-TEV-ATG5(1-275)	
VirGK3/ IcsB coexpression	pLS006/ pLS012	Strep-TEV-VirG(53-758)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK3/ ATG5 coexpression	pLS024/ pLS003	GST-Thr-TEV-VirG(53-758)/ Strep-TEV-ATG5(1-275)	
VirGK3/ IcsB coexpression	pLS024/ pLS013	GST-Thr-TEV-VirG(53-758)/ Strep-TEV-IcsB(1-494) IpgA(1-129)	

VirGK4/ ATG5 coexpression	pLS007/ pLS001	Strep-TEV-VirG(53-779)/ His-TEV-ATG5(1-275)	
VirGK4/ IcsB coexpression	pLS007/ pLS012	Strep-TEV-VirG(53-779)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK4/ ATG5 coexpression	pLS025/ pLS003	GST-Thr-TEV-VirG(53-779)/ Strep-TEV-ATG5(1-275)	
VirGK4/ IcsB coexpression	pLS025/ pLS013	GST-Thr-TEV-VirG(53-779)/ Strep-TEV-IcsB(1-494) IpgA(1-129)	
VirGK5/ ATG5 coexpression	pLS008/ pLS001	Strep-TEV-VirG(320-433)/ His-TEV-ATG5(1-275)	
VirGK5/ IcsB coexpression	pLS008/ pLS012	Strep-TEV-VirG(320-433)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK5/ ATG5 coexpression	pLS026/ pLS003	GST-Thr-TEV-VirG(320-433)/ Strep-TEV-ATG5(1-275)	
VirGK6/ ATG5 coexpression	pLS009/ pLS001	Strep-TEV-VirG(320-592)/ His-TEV-ATG5(1-275)	
VirGK6/ IcsB coexpression	pLS009/ pLS012	Strep-TEV-VirG(320-592)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK6/ IcsB coexpression	pLS027/ pLS013	GST-Thr-TEV-VirG(320-592)/ Strep-TEV-IcsB(1-494) IpgA(1-129)	
VirGK7/ ATG5 coexpression	pLS010/ pLS001	Strep-TEV-VirG(320-758)/ His-TEV-ATG5(1-275)	
VirGK7/ IcsB coexpression	pLS010/ pLS012	Strep-TEV-VirG(320-758)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK8/ ATG5 coexpression	pLS011/ pLS001	Strep-TEV-VirG(320-779)/ His-TEV-ATG5(1-275)	
VirGK8/ IcsB coexpression	pLS011/ pLS012	Strep-TEV-VirG(320-779)/ His-TEV-IcsB(1-494) IpgA(1-129)	
VirGK8/ ATG5 coexpression	pLS029/ pLS003	GST-Thr-TEV-VirG(320-779)/ Strep-TEV-ATG5(1-275)	
ESPRIT	pLS065	His-TEV-VirG(1-758)-BAP	*Nsi*I out
ESPRIT	pLS066	His-TEV-VirG(1-592)-BAP	*Nsi*I out
VirGE4	VirGE4	His-TEV-VirG(433-592)-BAP	

VirGE5	VirGE5	His-TEV-VirG(377-592)-BAP	
VirGE8	VirGE8	His-TEV-VirG(425-592)-BAP	
VirGE14	VirGE14	His-TEV-VirG(350-592)-BAP	
VirGE14	pLS099	His-TEV-VirG(350-592)	
VirGE15	VirGE15	His-TEV-VirG(341-592)-BAP	
VirGE16	VirGE16	His-TEV-VirG(363-592)-BAP	
VirGE19	VirGE19	His-TEV-VirG(361-592)-BAP	
VirGE19	pLS100	His-TEV-VirG(361-592)	
VirGE20	VirGE20	His-TEV-VirG(409-592)-BAP	
VirGE21	VirGE21	His-TEV-VirG(364-592)-BAP	
VirGE21	pLS101	His-TEV-VirG(364-592)	
VirGE24	VirGE24	His-TEV-VirG(430-592)-BAP	
VirGE33	VirGE33	His-TEV-VirG(353-592)-BAP	
VirGE57	VirGE57	His-TEV-VirG(469-758)-BAP	
VirGE57	pLS102	His-TEV-VirG(469-758)	
VirGE71	VirGE71	His-TEV-VirG(577-758)-BAP	
VirGE73	VirGE73	His-TEV-VirG(77-758)/VirG(638-758)-BAP	
VirGE76	VirGE76	His-TEV-VirG(340-758)-BAP	
VirGE76	pLS123	His-TEV-VirG(340-758)	
VirGE81	VirGE81	His-TEV-VirG(303-758)-BAP	
VirGE82	VirGE82	His-TEV-VirG(526-758)-BAP	
VirGE82	pLS103	His-TEV-VirG(526-758)	
VirGE94	VirGE94	His-TEV-VirG(419-758)-BAP	
VirGE94	pLS104	His-TEV-VirG(419-758)	

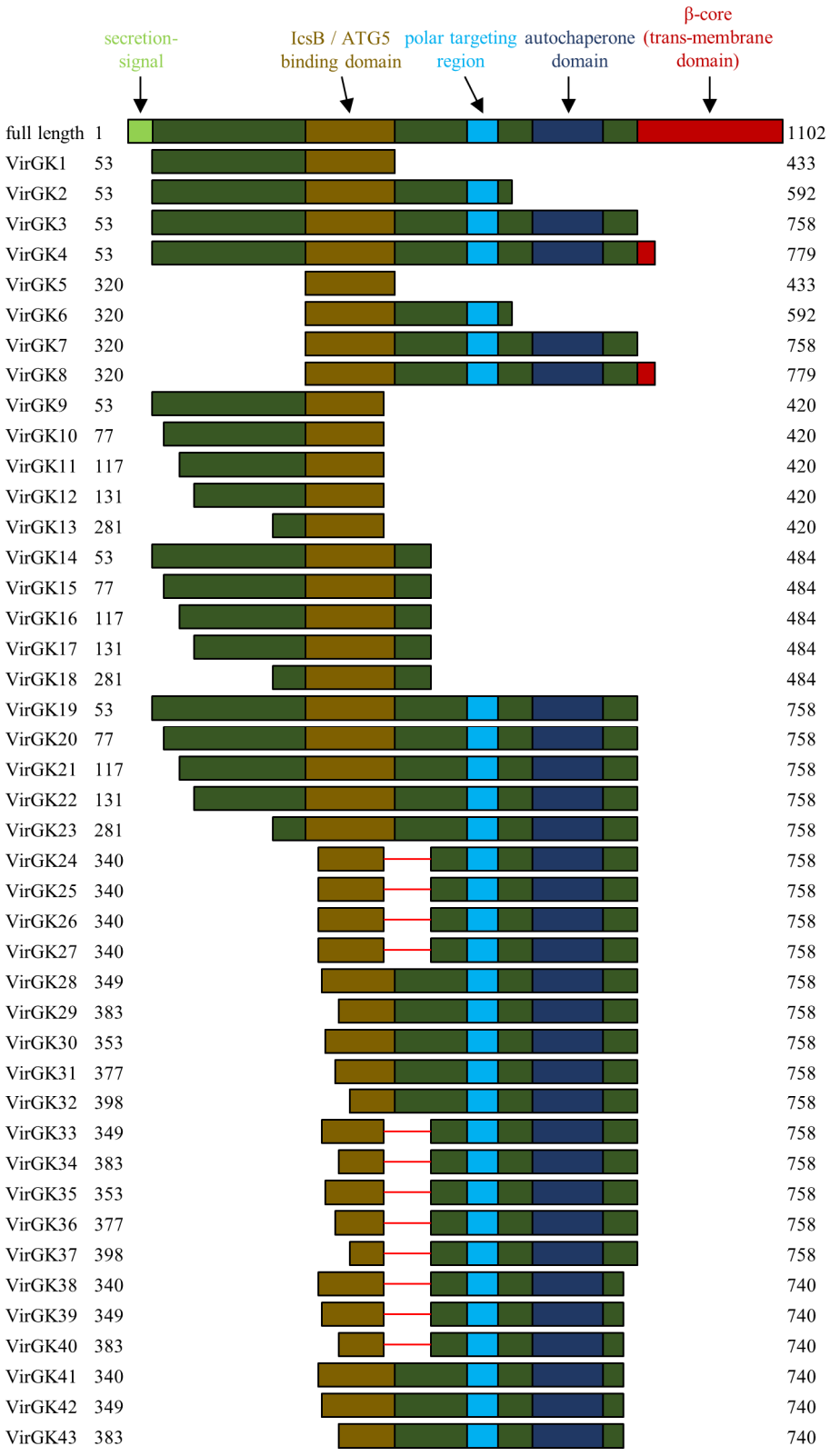


Figure 129: VirG protein constructs generated in this work.

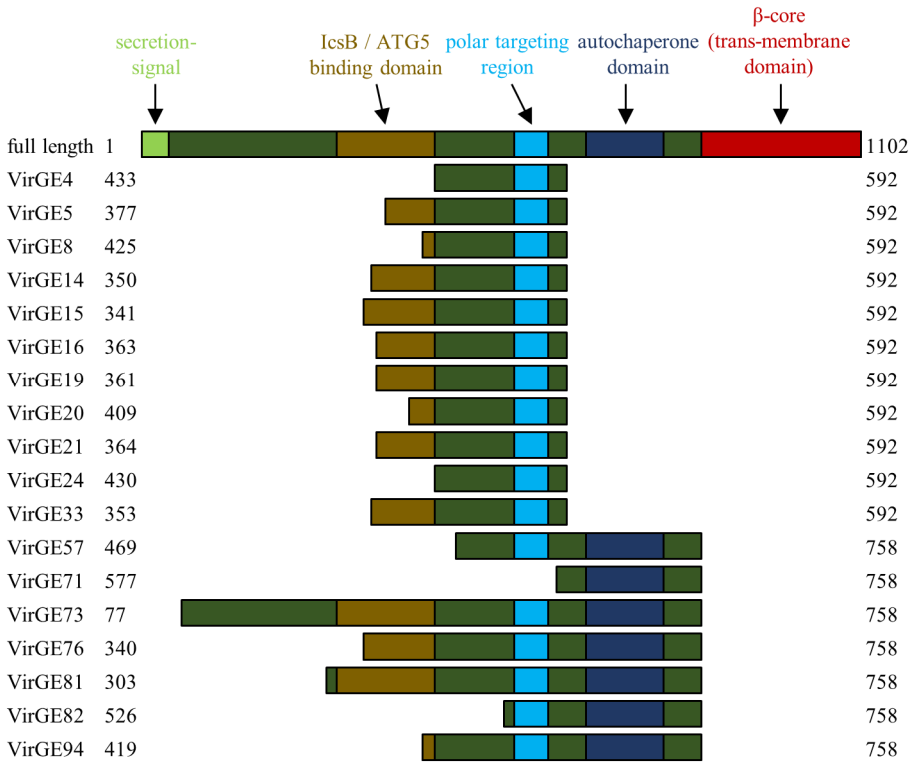


Figure 130: VirG protein constructs generated in this work, using the ESPRIT system.

Table 30: IcsB protein constructs generated in this work.

Description	Identifier	Construct	Mutations
IcsB	pLS012/030/031	His-TEV-IcsB(1-494) IpgA(1-129)	
IcsB	pLS013	Strep-TEV-IcsB(1-494) IpgA(1-129)	
IcsB	pLS105	His-TEV-IcsB(1-494)	
IcsB	pLS106	Flag-TEV-IcsB(1-494)	
IcsB	pLS107	Flag-TEV-IcsB(1-494) IpgA(1-129)	
IcsBK1	pLS032	His-TEV-IcsB(1-125) IpgA(1-129)	
IcsBK2	pLS033	His-TEV-IcsB(16-125) IpgA(1-129)	
IcsBK3	pLS034	His-TEV-IcsB(26-125) IpgA(1-129)	
IcsBK3	pLS209	GST-Thr-TEV-IcsB(26-125)	
IcsBK4	pLS035	His-TEV-IcsB(1-229) IpgA(1-129)	
IcsBK5	pLS036	His-TEV-IcsB(16-229) IpgA(1-129)	
IcsBK6	pLS037	His-TEV-IcsB(26-229) IpgA(1-129)	
IcsBK7	pLS038	His-TEV-IcsB(16-494) IpgA(1-129)	
IcsBK8	pLS039	His-TEV-IcsB(26-494) IpgA(1-129)	
IcsBK9	pLS040	His-TEV-IcsB(169-494) IpgA(1-129)	
IcsBK10	pLS041	His-TEV-IcsB(185-494) IpgA(1-129)	
IcsBK11	pLS042	His-TEV-IcsB(265-494) IpgA(1-129)	
IcsBK12	pLS076	His-TEV-IcsB(8-438) IpgA(1-129)	
IcsBK13	pLS077	His-TEV-IcsB(6-391) IpgA(1-129)	
IcsBK14	pLS078	His-TEV-IcsB(6-428) IpgA(1-129)	
IcsBK15	pLS079	His-TEV-IcsB(6-464) IpgA(1-129)	
IcsBK16	pLS080	His-TEV-IcsB(6-493) IpgA(1-129)	

IcsBK17	pLS081	His-TEV-IcsB(188-391) IpgA(1-129)	
IcsBK18	pLS082	His-TEV-IcsB(188-428) IpgA(1-129)	
IcsBK19	pLS083	His-TEV-IcsB(188-464) IpgA(1-129)	
IcsBK20	pLS084	His-TEV-IcsB(188-493) IpgA(1-129)	
IcsBK21	pLS085	His-TEV-IcsB(233-391) IpgA(1-129)	
IcsBK22	pLS086	His-TEV-IcsB(233-428) IpgA(1-129)	
IcsBK23	pLS087	His-TEV-IcsB(233-464) IpgA(1-129)	
IcsBK24	pLS088	His-TEV-IcsB(233-493) IpgA(1-129)	
IcsBK25	pLS089	His-TEV-IcsB(6-167) IpgA(1-129)	
IcsBK25	pLS096	Strep-TEV-IcsB(6-167)	
IcsBK26	pLS090	His-TEV-IcsB(6-176) IpgA(1-129)	
IcsBK26	pLS097	Strep-TEV-IcsB(6-176)	
IcsBK27	pLS091	His-TEV-IcsB(6-183) IpgA(1-129)	
IcsBK27	pLS098	Strep-TEV-IcsB(6-183)	
IcsBK28	pLS092	His-TEV-IcsB(177-493) IpgA(1-129)	
IcsBK29	pLS093	His-TEV-IcsB(214-493) IpgA(1-129)	
IcsBK30	pLS116	GST-Thr-TEV-IcsB(104-130)	
IcsBK30	pLS117	ZZ-His-TEV-IcsB(104-130)	
IcsBK31	pLS118	GST-Thr-TEV-IcsB(99-136)	
IcsBK31	pLS119	ZZ-His-TEV-IcsB(99-136)	
IcsBK32	pLS208	His-TEV-IcsB(121-494)	
IcsBSER1	pLS053	His-TEV-IcsB(1-494) IpgA(1-129)	Q237A, K238A, K239A
IcsBSER2	pLS054	His-TEV-IcsB(1-494) IpgA(1-129)	K447A, K448A, E449A
IcsBSER3	pLS055	His-TEV-IcsB(1-494) IpgA(1-129)	K285A, K286A, K287A
IcsBSER4	pLS056	His-TEV-IcsB(1-494) IpgA(1-129)	E81A, K82A, E85A
IcsBSER5	pLS057	His-TEV-IcsB(1-494) IpgA(1-129)	E182A, K183A
IcsBSER6	pLS058	His-TEV-IcsB(1-494) IpgA(1-129)	K228A, K229A
ESPRIT	pLS067	Flag-TEV-IcsB(1-494)	

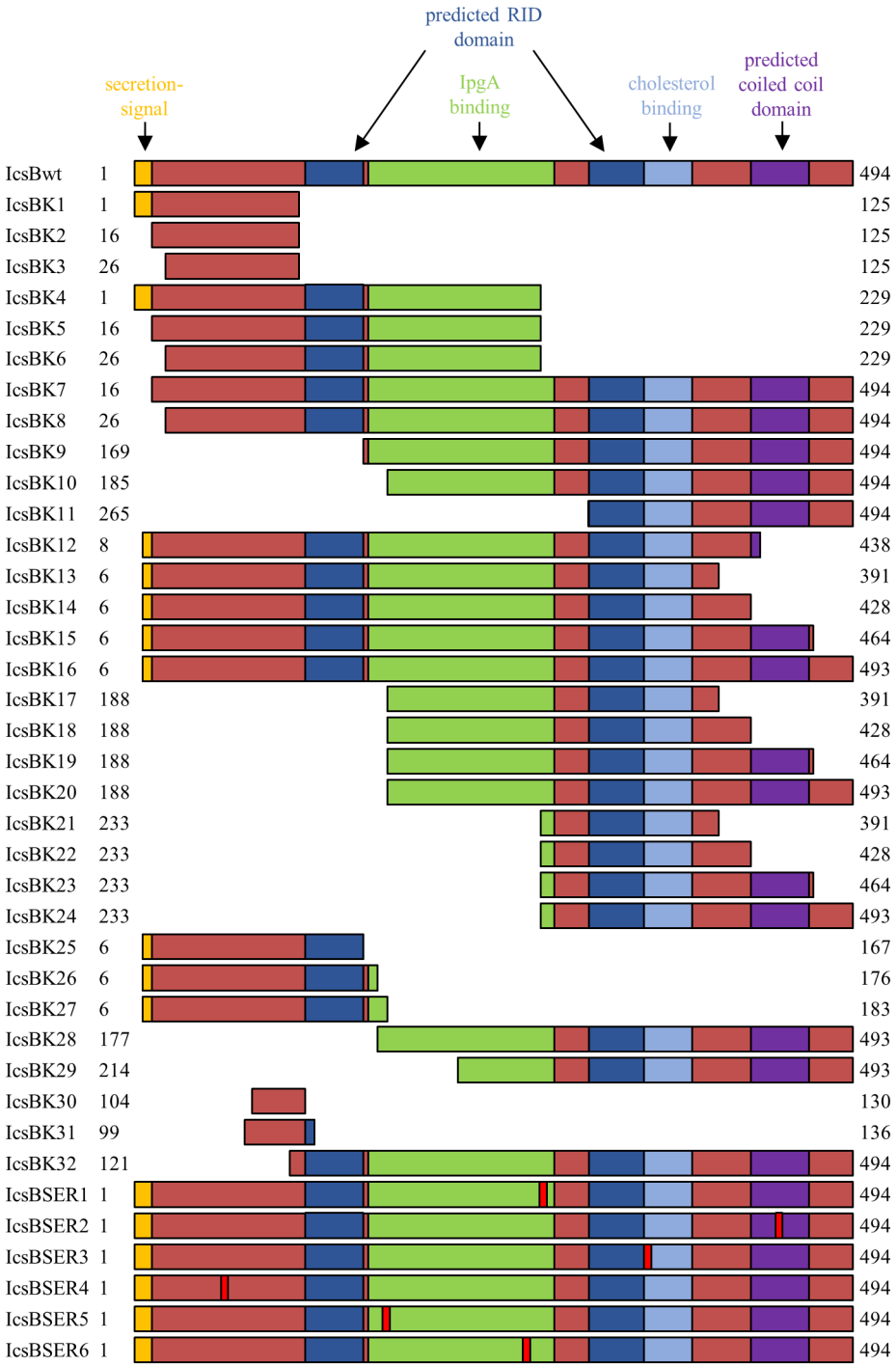


Figure 131: IcsB protein constructs generated in this work.

Table 31: ATG5 protein constructs generated in this work.

Description	Identifier	Construct	Mutations
ATG5	pLS001	His-TEV-ATG5(1-275)	
ATG5	pLS003	Strep-TEV-ATG5(1-275)	
ATG5 T75E	pLS059	His-TEV-ATG5(1-275)	T75E
ATG5 T75A	pLS060	His-TEV-ATG5(1-275)	T75A, T101T (silent)

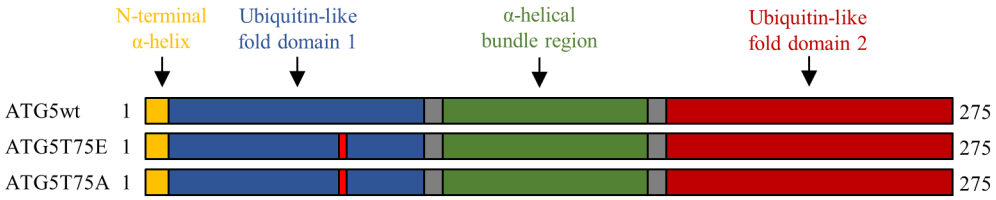


Figure 132: ATG5 protein constructs generated in this work.

Table 32: Other protein constructs generated in this work.

Description	Identifier	Construct	Mutations
ATG16L1	pLS070	His-MBP-TEV-Atg16L1(1-43)	
ATG16L1	pLS072	His-TEV-Atg16L1(1-43)	
ATG16L1	pLS074	GST-Thr-TEV-Atg16L1(1-43)	
ATG16L1	pLS069	His-MBP-TEV-Atg16L1(1-43)	44C
ATG16L1	pLS071	His-TEV-Atg16L1(1-43)	44C
ATG16L1	pLS073	GST-Thr-TEV-Atg16L1(1-43)	44C
WASP-GBD	pLS110	His-TEV-WASP(242-310)	
WASP-GBD	pLS111	GST-Thr-TEV-WASP(242-310)	
WASP-C	pLS114	GST-Thr-TEV-WASP(461-492)	
WASP-C	pLS115	ZZ-His-TEV-WASP(461-492)	
N-WASP-GBD	pLS108	His-TEV-N-WASP(207-275)	
N-WASP-GBD	pLS109	GST-Thr-TEV-N-WASP(207-275)	
N-WASP-GBD	pLS142	Strep-TEV-N-WASP(207-275)	
N-WASP-C	pLS112	GST-Thr-TEV-N-WASP(462-493)	
N-WASP-C	pLS113	ZZ-His-TEV-N-WASP(462-493)	
N-WASP-C	pLS143	Strep-TEV-N-WASP(462-493)	
N-WASP-VVCA	pLS147	GST-Thr-TEV-N-WASP(399-505)	
N-WASP-VVCA	pLS148	ZZ-His-TEV-N-WASP(399-505)	
N-WASP-VV	pLS149	GST-Thr-TEV-N-WASP(399-460)	
N-WASP-VV	pLS150	ZZ-His-TEV-N-WASP(399-460)	
Toca-1K1	pLS151	GST-Thr-TEV-Toca-1(389-480)	
Toca-1K2	pLS152	GST-Thr-TEV-Toca-1(389-488)	
Toca-1K3	pLS153	GST-Thr-TEV-Toca-1(383-480)	
Toca-1K4	pLS154	GST-Thr-TEV-Toca-1(383-488)	
Toca-1K5	pLS181	GST-Thr-TEV-Toca-1(1-293)	
Toca-1K6	pLS182	GST-Thr-TEV-Toca-1(245-373)	
Toca-1K7	pLS183	GST-Thr-TEV-Toca-1(245-477)	
Cip4K1	pLS155	GST-Thr-TEV-Cip4(392-480)	
Cdc42	pLS156	GST-Thr-TEV-Cdc42(1-178)	
Diaph2-GBD	pLS158	GST-Thr-TEV-Diaph2(88-476)	S149 → SIVGSKVT (isoform 2/3)
Diaph3-GBD	pLS159	GST-Thr-TEV-Diaph3(264-488)	Δ1-263 (isoform 1)
Daam2-GBD	pLS160	GST-Thr-TEV-Daam2(39-427)	
Fhod1-GBD	pLS161	GST-Thr-TEV-Fhod1(44-468)	
Fmnl2-GBD	pLS163	GST-Thr-TEV-Fmnl2(16-487)	
Fmnl3-GBD	pLS164	GST-Thr-TEV-Fmnl3(16-477)	

N-WASP-GBD–N-WASP-C	pLS144	His-TEV-N-WASP(207-275)-GGS-N-WASP(462-493)	
N-WASP-GBD–IcsBK30	pLS145	His-TEV-N-WASP(207-275)-GGS-IcsB(104-130)	
N-WASP-GBD–IcsBK31	pLS146	His-TEV-N-WASP(207-275)-GGS-IcsB(99-136)	

Danksagung

Zuerst möchte mich bei meinem Doktorvater Prof. Dr. Dieter Jahn, für die Betreuung meiner Doktorarbeit als Mentor bedanken. Weiterhin gilt mein Dank Prof Dr. Wulf Blankenfeldt und Prof. Dr. Michael Hust für die Übernahme des Koreferats und des Prüfungsvorsitzes.

Mein herzlichster Dank gilt Andrea Scrima für die Möglichkeit dieses interessante Thema zu bearbeiten. In seiner Nachwuchsgruppe herrschten immer eine gute Stimmung und ein angenehmes Arbeitsklima. Außerdem hat er einen erstklassigen Job bei der Betreuung aller seiner Doktoranden gemacht. Seine Bürotür stand für uns immer offen, wenn wir Probleme hatten oder Ratschläge brauchten. Als sein erster Doktorand war es eine gute Erfahrung für mich den Aufbau einer Juniorarbeitsgruppe mitzuerleben.

Außerdem möchte ich meinen Kollegen aus der Strukturbiologie der Autophagie: Archna, Milica Bajagic, Caroline Behrens, Pooja Sadana, Stefan Schmelz und nicht zuletzt Petra Völler danken für die schöne Zeit und die Hilfe bei den vielen kleinen und größeren Problemen. Meinen Studenten Pascal Bartling und Daniela Kuhn möchte ich danken für den Beitrag, den sie zu meiner Arbeit beigesteuert haben.

Bei den vielen anderen Kollegen aus der Strukturbiologie und der Rekombinanten Proteinexpression, insbesondere bei Jörn Krauße, Peer Lukat und Johannes Spehr, möchte ich mich bedanken für die vielfältige Hilfe und die interessanten Diskussionen.

Dr. Darren Hart und Phillipe Mas möchte ich danken für die Möglichkeit das ESPRIT System zu nutzen und für die exzellente Betreuung vor Ort. Durch die Ergebnisse dieses Systems sind große Teile dieser Arbeit erst möglich geworden.

Mein herzlichster Dank gilt außerdem meiner Freundin, meiner Familie und all meinen Freunden, die mich während der Doktorarbeit begleitet und die viel zu den schönen Jahren in Braunschweig und zu meiner Motivation beigetragen haben.